THE UNIVERSITY
of EDINBURGH

Investigating the linkage between *Trypanosoma brucei* pleomorphism and antigen switch frequency

Kirsty Ross McWilliam

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of

Doctor of Philosophy

University of Edinburgh

School of Biological Sciences

May 2019

Declaration

I declare that the content of this thesis is my own work and that all contributions and collaborations have been explicitly acknowledged in the text. No material presented in thesis has been submitted to any other university or for any other degree.

Kirsty McWilliam

Lay Summary

Trypanosoma brucei is a single-celled parasite which lives in the blood of its mammalian host and is spread by the tsetse fly. *T. brucei* is able to survive in the bloodstream, despite immune attack, by periodically switching its surface protein coat (called a Variable Surface Glycoprotein, VSG, coat) to a new form which is unrecognisable to the immune system. A second strategy, called differentiation, also prolongs infection. When high numbers of dividing 'slender' parasites accumulate in the blood, the parasites differentiate to non-dividing 'stumpy' forms which prevents rapid host death. These stumpy forms are important for the spread of disease because they are the only form that can survive being ingested by a tsetse fly. When trypanosomes are grown for long periods in the laboratory they lose the ability to differentiate from slender to stumpy forms. It has also been reported that these laboratory-adapted trypanosomes switch their VSG surface coats significantly less than differentiation-competent populations of trypanosomes. It therefore appears that differentiation capacity and VSG switch frequency may be linked. To investigate if these processes are coupled, I studied trypanosomes which were differentiation competent but in which different genes responsible for making stumpy forms could be silenced. This provided a tool to study VSG switch rate in the same populations where differentiation capacity could be switched 'on' or, upon silencing of a stumpy formation gene, 'off'.

Two different experimental approaches were used to ask if VSG switch rate is directly linked to differentiation capacity. The first approach measured the VSG switch rate of differentiation competent and incompetent trypanosomes, and found that there was no significant difference in VSG switch rate between the two populations. The second approach studied the number and identity of VSGs within the two populations. Differentiation incompetent populations did not express less, or different, VSGs. This demonstrated that VSG switch rate and differentiation capacity are not linked. To back up our observations, a population of differentiation-competent trypanosomes was grown in the laboratory until the parasites lost the capacity to differentiate. When the VSG switch rate of the differentiation-incompetent population was compared to that of the starting population, the VSG

switch rate had not decreased. This provided further evidence that the VSG switch rate and differentiation capacity are not linked. To understand why the parasites had lost the capacity to differentiate, I studied the genes which were either expressed more or less of after laboratory-adaptation. This analysis identified a number of genes that could be novel regulators of slender to stumpy form differentiation.

Abstract

African trypanosome infections are characterised by antigenic variation to avoid host immunity and by the production of transmission stages to promote disease spread. Laboratory-adapted ('monomorphic') lines of *Trypanosoma brucei* are reported to switch their expressed VSG antigenic coat at a much lower frequency than 'pleomorphic' populations recently transmitted through tsetse flies. These laboratory-adapted parasites also lose the capacity to differentiate into transmission competent 'stumpy forms'. It is unclear if the reduced rate of antigenic variation is directly coupled to the loss of pleomorphism or whether the processes, although co-selected by multiple passage, are independent.

To address whether monomorphism caused a concomitant change in the frequency of antigenic variation, an 'inducible monomorphism' model was exploited. This exploited pleomorphic RNAi cell lines that would inducibly silence genes required for stumpy formation. This provided a tool to switch pleomorphism 'on' or 'off' inducibly, without long term passage. Thereafter, two approaches were used to ask if the induction of monomorphism directly influenced antigenic variation: *in vitro* flow cytometry-based VSG switch assays, and VSGseq, a targeted sequencing approach. These assays demonstrated that the induction of monomorphism did not reduce VSG switch rate nor generate a reduction in expressed VSG diversity, or change the expressed VSG subset.

To extend this analysis further, the prolonged *in vitro* passage of a pleomorphic cell line was used to select isogenic monomorphic populations with reduced capacity to generate stumpy forms. These selected monomorphs cells did not exhibit a reduction in VSG switch rate compared to the parental pleomorphic population, thus corroborating the observations with the induced monomorphic cells. To understand the loss of pleomorphism in the selected cells, an 'evolve and re-sequence' approach and RNAseq analysis was adopted. Interestingly, the 'selected monomorphs' were depleted of transcripts whose expression was associated with the stumpy and insect forms of the parasite. Particularly, a number of CCCH zinc finger proteins, such as ZC3H20, were significantly downregulated. These RNA-binding proteins could represent novel regulators of slender to stumpy differentiation.

Overall our results demonstrate that *T. brucei* antigen switch frequency and pleomorphism can be uncoupled, and provide new insight into the molecular control of stumpy formation.

Acknowledgements

My thanks firstly go to Prof Keith Matthews for his supervision throughout the course of my PhD. Anything that could help further my progress he has made possible and his patience when faced with my mistakes, stubbornness or refusal to listen has been greatly appreciated. Thank you also to my second supervisor, Dr Liam Morrison, and postgraduate committee member, Dr Joanne Thompson, for their support and advice over the last three years.

The incredible people in the Matthews Lab, in particular Balazs, Mathieu and Ele, have made this experience an easier and far more enjoyable one. I thank everyone who has at some point assisted me technically in the lab, in particular Julie Young for the huge task of training me in animal work and Dr Eva Rico for her guidance when I began my work in the lab.

The following people have all contributed in some way to the data presented in this thesis and, in addition to the acknowledgements in the text, I wish to repeat those thanks here: Dr Alasdair Ivens for the data processing in Chapter 5; Dr Monica Mugnier for welcoming me into her lab and for the data processing performed in Chapter 4; Dr Martin Waterfall for his help in FACS setup; Dr Olivia Swann for providing me with her blank thesis format and finally, Dr Mathieu Cayla for the ZC3H20 data presented in Chapter 5.

This thesis was made possible by funding from the MRC and white chocolate and raspberry muffins from the Noreen and Kenneth Murray Library Café.

List of Abbreviations

4EIP	4E-interacting protein
8-pCPTcAMP	8-(4-Chlorophenylthio)adenosine 3′,5′-cyclic monophosphate
AAT	Animal African trypanosomiasis
AC	Adenylate cyclase
ARE	AU-rich element
ATF	Adipose tissue form
ATP	Adenosine triphosphate
BES	Bloodstream form expression site
BLA	Blasticidin resistance
BLAST	Basic local alignment search tool
bp	Base pair
BPGA	Biphosphoglyceric acid
BSF	Bloodstream form
cAMP	cyclic 3', 5'-adenosine monophosphate
CCA	Cis-aconitate
cDNA	Complementary DNA
CDS	Coding sequence
CNS	Central nervous system
CTD	C-terminal domain
DAPI	4′,6-diamidino-2-phenylindole
DEAE-C	Diethylaminoethyl cellulose
DHAP	Dihydroxyacetone phosphate

DMFO	Difluoromethylornithine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DND*i*	Drugs for Neglected Diseases *initiative*
DOX	Doxycycline
DSB	Double-stranded break
DYRK	Dual-specificity tyrosine phosphorylation-regulated kinase
E & R	Evolve and resequence
EATRO	East African Trypanosomiasis Research Organisation
eGFP	Enhanced green fluorescent protein
ELC	Expression-linked copy
EMA	European Medicines Agency
ES	Expression site
ESAG	Expression site associated gene
ESB	Expression site body
FACS	Fluorescence-activated cell sorting
FAZ	Flagellar attachment zone
FSC	Forward scatter
G3P	Glycerol 3-phosphate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Gene conversion
gDNA	Genomic DNA
GFP	Green fluorescent protein

GPI	Glycosylphosphatidylinositol
GRESAG	Genes related to ESAG4
HAT	Human African trypanosomiasis
HDL	High density lipoprotein
HSV-TK	Herpes simplex virus thymidine kinase
HYP2	Hypothetical protein 2
IFA	Immunofluorescence analysis
IM	Intramuscular
INDEL	Insertion/deletion
INT1/2	‘Intermediate 1/2'
IP	Intraperitoneal
ISG	Invariant surface glycoprotein
IV	Intravenous
K	Kinetoplast
kb	Kilobase
kDNA	Kinetoplast DNA
Log2FC	Log2(fold change)
MACS	Magnetic-activated cell sorting
MAP	Microtubule-associated protein
MAPK5	Mitogen-activated protein kinase 5
MCP	Mitochondrial carrier protein
mRNA	Messenger RNA
N	Nucleus

NCBI	National Centre for Biotechnology Information
NECT	Nifurtimox eflornithine combination therapy
NEK	NimA-related kinase
NIT1	Nitrilase-like protein 1
NPC	Nuclear pore complex
NTD	N-terminal domain
NUP-1	Nuclear peripheral protein 1
ORF	Open reading frame
OSF	Observed switching frequency
PAC	Puromycin N-acetyltransferase
PAD	Protein associated with differentiation
PCF	Procyclic form
PI	Propidium iodide
PIL	Personal licence
PKC	Protein kinase C
PP1	Protein phosphatase 1
PUR	Puromycin resistance
RACK1	Receptor for activated protein C kinase 1
RAGE	Rotating field gel electrophoresis
RBP	RNA binding protein
RITseq	RNA interference target sequencing
RNA	Ribonucleic acid
RNAi	RNA interference

RNAPI	RNA polymerase I
RNAseq	RNA sequencing
RPKM	Reads per kilobase of transcript, per million mapped reads
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RT	Reverse transcriptase
RT-PCR	Reverse transcription polymerase chain reaction
SBP	Serial blood passage
SD	Standard deviation
SEM	Standard error of the mean
SIF	Stumpy induction factor
SILAC	Stable isotope labelling with amino acids in cell culture
SL	Spliced leader
SNP	Single nucleotide polymorphism
SRA	Serum resistance associated
SS	Single stranded
SSC	Side scatter
TbTfR	Trypanosoma brucei transferrin receptor
TERT	Telomerase reverse transcriptase
TIM	Triosephosphate isomerase
TLF	Trypanosome lytic factor
TNF	Tumour necrosis factor
TREU	Trypanosomiasis Research Edinburgh University

TS	Trans-sialidase
UTR	Untranslated region
VAT	Variable antigen type
VSG	Variant surface glycoprotein
WGS	Whole genome sequencing
WHO	World Health Organization
WT	Wild type
ZC3H	CCCH zinc finger protein
ZFP	Zinc finger protein

For Granda

CONTENTS

1 : INTRODUCTION

1.1 *Trypanosoma brucei*

Kinetoplastid parasites are a fascinatingly complex and diverse group of eukaryotes. Their cell biology is replete with curious variations of the norm - novel mechanisms of gene expression control, metabolism and cell structure name only a few. A single, large mitochondrion characterises members of the Kinetoplastida order, within which is a complex network of kinetoplastid DNA (kDNA) comprised of mini- and maxi-circles (Cavalcanti and de Souza, 2018; Lukes et al., 2002). One such example of a parasite within this intriguing Order is the unicellular flagellate *Trypanosoma brucei*.

1.1.1 Human African trypanosomiasis (HAT)

The kinetoplastids are responsible for profound morbidity, mortality and economic loss, particularly in rural areas where health infrastructure is lacking or absent. *T. brucei* is the causative agent of the neglected tropical disease Human African Trypanosomiasis (HAT), otherwise known as sleeping sickness. Those at highest risk of exposure to the disease typically live in rural communities which rely on agriculture, animal husbandry, hunting or fishing (WHO, 2018). HAT is a highly focal disease, its geographic distribution strictly governed by the restriction of the trypanosome's arthropod vector, the tsetse fly (*Glossina* spp.), to 360 HAT foci within 36 Sub-Saharan African countries (Franco et al., 2014a; Wamwiri and Changasi, 2016) (Figure 1.1). The Palpalis group of *Glossina* (G. *palpalis palpalis, G. p. gambiensis* and *G. fuscipes*) are found in West and Central Africa and transmit the *T. b. gambiense* subspecies, whereas the Morsitans group (*G. morsitans*) are found in East Africa's savannah woodlands and are involved in *T. b. rhodesiense* transmission (Franco et al., 2014a; Wamwiri and Changasi, 2016).

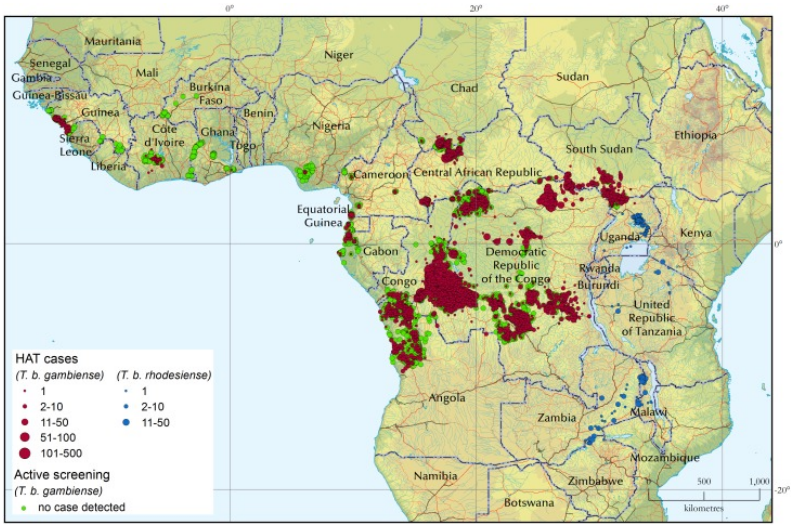


Figure 1.1 Distribution of *T. b. rhodesiense* and *T. b. gambiense* HAT cases from 2010-2014. Figure from (Franco et al., 2017).

Most African trypanosomes are readily lysed by human ApoL1, a high density lipoprotein (HDL)-associated serum protein. ApoL1 forms anionic pores within the trypanosome lysosome following the internalisation of trypanosome lytic factor (TLF) into the endocytic pathway by the trypanosome's haptoglobin-haemoglobin receptor (TbHpHbR) (Drain et al., 2001; Perez-Morga et al., 2005; Vanhamme et al., 2003). Both *T. b. gambiense* and *T. b. rhodesiense*, however, have evolved means to resist host trypanolytic serum factors, and thus cause human infection. An expression site related gene termed 'serum resistance associated' (SRA), identified in the ETat 1.10 expression site (ES), is responsible for conferring immunity to ApoL1 in serum-resistant *T. b. rhodesiense* strains (De Greef and Hamers, 1994; Xong et al., 1998). Unlike most ESAGs, SRA is not expressed on the surface, and is instead trafficked to the flagellar pocket before being endocytosed into the endolysosomal system (Stephens and Hajduk, 2011). It is here that SRA colocalises and binds to ApoL1, preventing it from forming pores within the lysosome. *T. b. gambiense* resistance appears to be categorised into two groups, 1 and 2 (Capewell et al., 2011; Uzureau et al., 2013). Group 1 parasites possess a conserved single amino acid substitution within their HpHb receptor which lessens the uptake of TLF-1 (DeJesus et al., 2013; Kieft et al., 2010; Uzureau et al., 2013)

and furthermore express *TgsGP*, a gene which encodes a 47kDa surface protein which localises to the flagellar pocket (Berberof et al., 2001; Uzureau et al., 2013). Deletion of *TgsGP* in wild-type group 1 *T. b. gambiense* was shown to render parasites sensitive to lytic human serum, and reintroduction of the gene into the knockout clones could restore resistance, thus demonstrating that TgsGP is essential for protection against human serum in group 1 *T. b. gambiense* (Capewell et al., 2013a). Group 2 parasites show variability in their resistance, and since they have still been shown to internalise TLF it has been proposed their resistance may be due to neutralisation of the effects of ApoL1, or by a different means of avoidance than Group 1 (Capewell et al., 2011).

Though the two subspecies may be morphologically indistinguishable, the diseases caused by *T. b. gambiense* and *T. b. rhodesiense* differ widely in their epidemiology, onset of symptoms and prognosis (Franco et al., 2014a). Infection with *T. b. gambiense* accounts for >98% of all reported cases and causes chronic infection (WHO, 2018). Those infected can expect to survive for months, or even years, before the onset of symptoms in latter stages of the disease (Checchi et al., 2008). In contrast, *T. b. rhodesiense* disease is fulminant with disease progression rapidly reaching fatality within the course of around 6 months (Odiit et al., 1997). Uganda is currently the only country known to house both forms of the disease, however these are separated geographically (WHO, 2018). Both diseases are characterised by an early (haemolymphatic) stage and a late (neurological or meningo-encephalic) stage. In the early stages of disease, the parasites replicate within the blood, lymph and subcutaneous tissue (Trindade et al., 2016). This stage of the disease may be accompanied by non-specific symptoms such as malaise, headaches, intermittent fevers and dermatological problems, such as pruritus (Franco et al., 2014a; Kennedy, 2013). As the infection progresses, the parasites cross the blood brain barrier and infiltrate the central nervous system (CNS). In this latter stage of disease symptoms intensify, with sensory, motor and behavioural disturbances occurring alongside a loss of coordination and the affliction which gives the disease its name, a disruption in sleeping patterns (Kennedy, 2013). Left untreated HAT is usually fatal, though it is becoming increasingly apparent that is some individuals infected with *T. b. gambiense* are able to progress to a seemingly

asymptomatic state, or even resolve infection entirely, in the absence of treatment (Capewell et al., 2016; Jamonneau et al., 2012; Sudarshi et al., 2014). In a 2 year study of cytokine levels in *T. b. gambiense* HAT patients, trypanotolerant individuals (who had positive serological responses but undetectable parasitaemia under the microscope) were associated with higher levels of inflammatory cytokines, in particular IL8, compared to HAT patients and healthy endemic controls (Ilboudo et al., 2014). The prevalence of asymptomatic carriers could threaten WHOs disease elimination programme by accommodating continued disease transmission between humans. Thus, more must be done to understand the impact of these individuals on disease transmission and elimination (Berthier et al., 2016).

1.1.2 Animal African trypanosomiasis (AAT)

The *T. b. rhodesiense* subspecies, as well as *T. b. brucei*, is capable of infecting both wild and domestic animals such as cattle, wildebeest, horses and dogs, though the animal disease burden caused by these subspecies is considerably less than that caused by *T. congolense* or *T. vivax* (Giordani et al., 2016). The resulting wasting disease, termed nagana, is characterised by anaemia, loss of productivity and fertility and emaciation. In areas which rely so heavily on agriculture and animal husbandry, AAT limits the potential for agricultural development drastically. Animal mortality, reduced crop cultivation and diminished meat and milk yields all impact farmers' livelihoods. Indeed, AAT is considered the most economically significant disease of livestock in Africa, losing the industry billions of US dollars in revenue annually (Giordani et al., 2016; Welburn et al., 2006).

1.1.3 Current therapeutic options

The World Health Organization (WHO) have targeted elimination of *T. b. gambiense* sleeping sickness as a public health problem by the year 2020, and sustainable disease elimination by 2030 (WHO, 2012). National Control Programmes' implementation of surveillance, control and treatment guidelines set out by WHO and the creation of public-private partnerships to guarantee the free

supply of therapeutic drugs, has allowed the number of reported cases to drop to 2804 in 2015, the lowest number of reported cases since global data collection began ~80 years ago (WHO, 2012, 2018) (Figure 1.2). However, under-detection and under-reporting, due to an unstable social climate or geographical inaccessibility, means that the likely number of cases is in reality nearer 20,000 (WHO, 2013). Elimination of *T. b. rhodesiense* is considered less feasible since it causes a zoonotic disease, the parasite being predominantly found within wild and domestic animal reservoirs rather than human hosts, these being harder to regulate (Franco et al., 2014b). Though *T. b. gambiense* causes anthroponotic disease, such is our limited knowledge of *T. b. gambiense* animal reservoirs it is worried that cryptic animal vectors we do not yet know about could compromise elimination efforts (Buscher et al., 2018).

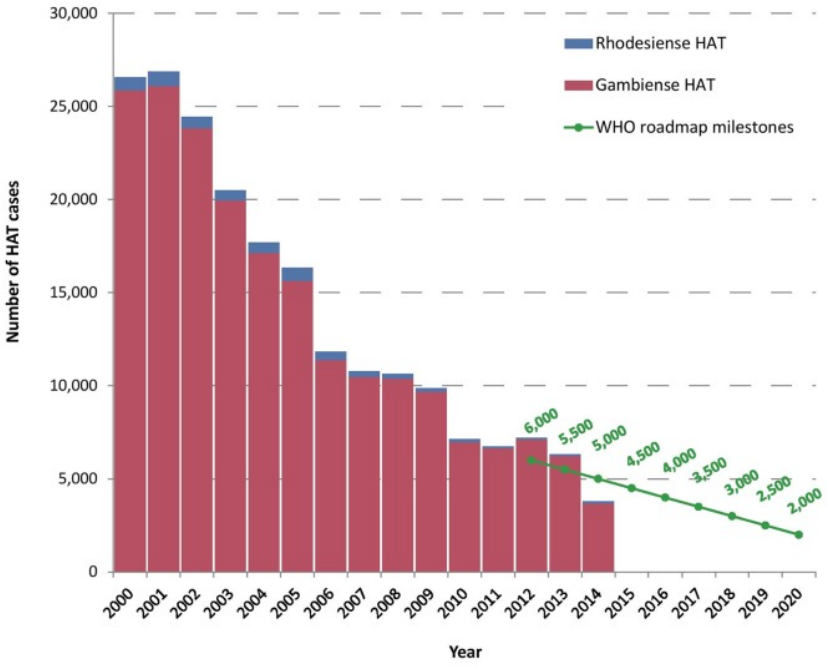


Figure 1.2 Number of recorded *T. b. gambiense* (red) and *T. b. rhodesiense* (blue) HAT cases per year from 2000-2014. The numbers in green are the projected cases per year as set out in WHOs elimination milestones. Figure from (Franco et al., 2017).

Until very recently, five different drugs have been used to treat HAT (4 of which are licensed), however none of these drugs is without flaw (Table 1). High toxicity, long treatment periods, parenteral administration, drug resistance and low efficacy all contribute to the unsatisfactory nature of these chemotherapeutic options (Barrett et al., 2007). On the 16th November 2018, however, the European Medicines Agency (EMA) adopted a positive scientific opinion of fexinidazole, the only available oral monotherapy to treat both early and late *T. b. gambiense* HAT to date, following a successful Phase 2/3 trial (Deeks, 2019; Mesu et al., 2018). The drug is taken orally once-a-day for 10 days. Fexinidazole was initially discovered in the late 1970's by Hoechst AG (now Sanofi), however it was not until 2009, when the non-profit research and development organisation Drugs for Neglected Diseases *initiative* (DND*i*) entered into a partnership with Sanofi, that the drug was entered into clinical development as a cure for both early and late stages of *T. b. gambiense* HAT (Pollastri, 2018). Fexinidazole, a DNA synthesis inhibitor, was shown to cure murine models of acute *T. b. gambiense* and chronic *T. b. brucei* (the model for stage 2 HAT) (Torreele et al., 2010) and in a randomised phase 2/3 open-label non-inferiority trial with 394 patients from the Democratic Republic of Congo and the Central African Republic was shown to have 91% success at 18 months (success being deemed as patient survival, no evidence of trypanosomes in any bodily fluid, no rescue medication and having a cerebrospinal fluid white blood cell count ≤20µl). This was comparable to 98% success at 18 months observed upon treatment with nifurtimox eflornithine combination therapy (NECT) and therefore within the margin of acceptable difference (which was set at 13%) (Mesu et al., 2018). Furthermore, unlike other late stage HAT drugs, fexinidazole displays little to no non-specific cytotoxicity. Fexinidazole is therefore an exciting development in the race to eliminate *T. b. gambiense* HAT as a public health problem by 2020 since it eliminates the need for patient hospitalisation and intravenous drug treatment. A fexinidazole phase 3b trial that started in 2016, and has enrolled 116 patients, is due for completion in March 2020 (DND*i*, 2019b).

The benzoxaborole, acoziborole (AN5568/SCYX-7158) is another promising drug candidate which is currently being tested in Phase 2/3 trials for the treatment of

stage 2 *T. b. gambiense* HAT. Acoziborole is the result of the DND*i*'s own lead optimisation optimisation project (having initially been identified as a hit in an Anacor chemical library screen) and was selected as a pre-clinical candidate in 2009 (DND*i*, 2019a). The therapeutic dose of the drug is administered once, orally, as three tablets (DND*i*, 2019a). The phase 2/3 trial commenced in the last quarter of 2016 in the Democratic Republic of Congo and as of 2018, 191 patients were enrolled. The trial is expected to be completed by April 2020, with approval from the EMA to be sought in 2021 (DND*i*, 2019a). The target of acoziborole has been chemically and genetically validated as the endonuclease CPSF3 (*Tb*927.4.1340) and the mode of action as inhibition of mRNA maturation (Wall et al., 2018).

To date, no functional vaccine to protect against HAT has been developed. The transferrin receptor, invariant surface glycoproteins (ISGs) 64, 65 and 100, β-tubulin, actin, microtubule-associated proteins (MAPs), trans-sialidases, cation pumps, GPI and cysteine proteases have all been quoted as being 'promising' vaccine candidates yet not one of these has led to a successful field trial (reviewed in (La Greca and Magez, 2011; Magez et al., 2010)). Two factors have been blamed for hampering vaccine development efforts. Firstly, the parasite is able to switch its Variant Surface Glycoprotein (VSG) surface coat to an antigenically distinct form and remove VSG-bound antibodies from its surface (described in detail in Section 1.2) and secondly, trypanosome infection has been shown to abolish host B-cell memory function, therefore preventing proper induction of a protective response (Radwanska et al., 2008).

The application of innovative technology could offer novel methods of trypanosome control. Nanobodies are ~15kDa single-domain antibodies derived from camelids which can be used to efficiently target a drug to parasites (Stijlemans et al., 2017). Nanoparticles loaded with pentamidine and coupled to a nanobody were shown to reduce the curative dose for *T. b. gambiense* in a mouse model by 100-fold and, furthermore, circumvent known drug resistance mechanisms (Stijlemans et al., 2017; Unciti-Broceta et al., 2015). Silver and gold nanoparticles were recently shown to inhibit *T. b. brucei* growth by >90% *in vitro* with 200-fold more selectivity for the

parasites than mammalian cells (Adeyemi et al., 2018). However, nanoparticle treatment did not decrease parasite burden *in vivo*.

Table 1 Current HAT drug treatment options. Data compiled from (Babokhov et al., 2013; Barrett et al., 2007; Eperon et al., 2014; Fairlamb and Horn, 2018; Kennedy, 2013; Wiedemar et al., 2018).

DRUG	SUBSPECIES TREATED	STAGE TREATED	ADMINISTRATION	SIDE EFFECTS	MODE OF ACTION	RESISTANCE IN FIELD
PENTAMIDINE	*T. b. gambiense* only	Early	IM injection. 4mg/kilo daily for 7-10 injections.	Generally well tolerated.	Selective accumulation within trypanosome cell. Uptake by P2/HAPT1/LAPT1 transporters. Mitochondrial processes targeted.	Six isolates from relapsed patients found to carry mutations in their P2 transporter. 40-50 fold less sensitive
SURAMIN	*T. b. rhodesiense* only	Early	Slow IV injection. 5 injections of 20mg/kilo every 3-7 days over 4 weeks.	Neuropathy, renal failure, skin lesions and anaemia common.	Not clear. Uptake by endocytosis. Thought to bind to glycolytic enzymes to disrupt glycolysis and may also disrupt pentose phosphate pathway.	No reports
MELARSOPROL	Both	Late	IV injection. 2.2mg/kilo daily for 10 days. Must be adminitered from glass syringes as the compound dissolves plastic syringes.	Severe. Painful injection. 5-10% of patients will develop post-treatment reactive encephalopathy, of which half will die.	Not clear. Selective uptake by AT1 and AQP2. Thought to competitively inhibit trypanothione reductase.	Widespread treatment failure of 20-50% in some areas. Defective *AQP2* gene responsible.
EFLORNITHINE	*T. b. gambiense* only	Late	IV infusion. 100mg/kilo every 6 hours for 14 days.	Fever, peripheral neuropathy, GI problems and bone marrow toxicity.	Carrier mediated uptake. Inhibitor of ornithine decarboxylase.	No reports
NIFURTIMOX	*T. b. gambiense* only	Both	Given in combination with eflornithine as IV infusion. 200mg/kg every 12hours for 7 days of eflornithine. 15mg/kg for 10 days of nifurtimox.	Anorexia, vomiting/diarrhoea.	Reduction of prodrug forms a toxic reactive nitrile.	No reports

1.1.4 Trypanosome morphology and cell cycle

T. brucei is a unicellular parasite which must undergo extreme morphological rearrangement throughout the different stages of its lifecycle. The cell's elongated shape, tapering to the anterior, is strictly maintained by its highly organised cytoskeleton directly below the pellicular membrane (Portman and Gull, 2012). The cortical microtubules are highly polarised, with their + ends all orientated towards the posterior of the cell (Robinson et al., 1995), and are cross-linked to both each other and the flagellar membrane (Gull, 1999). In addition to their role in maintaining cellular form, the network of microtubules furthermore engage in fundamental roles in mitosis, cytokinesis, segregation and organelle positioning (Robinson et al., 1995). The organelles are focused within the posterior end of the cell with the most posterior structure being the flagellar pocket, the only site of endo-/exocytosis within the cell (Field and Carrington, 2009) (Figure 1.3). A single flagellum extends from the flagellar pocket to just beyond the anterior pole of the cell, and is attached to the entire length of the cell body by a structure termed the flagellar attachment zone (FAZ). Other single copy organelles include the mitochondrion (which uniquely reaches along the length of the cell), kinetoplast, lysosome and basal body.

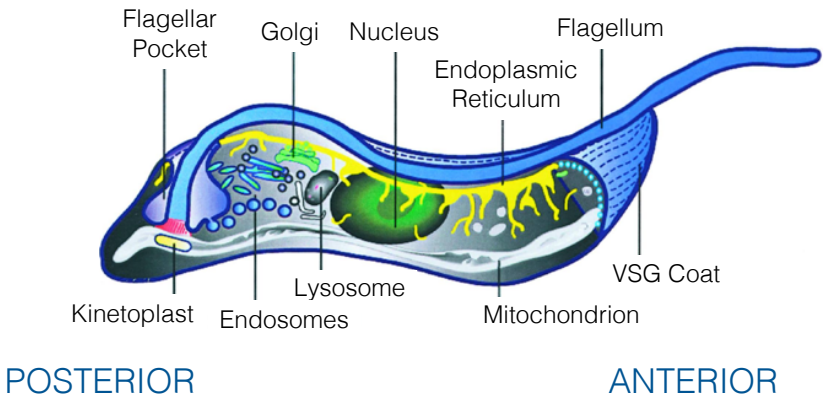


Figure 1.3 *T. brucei* cell structure. The locations of various organelles are highlighted. Figure adapted from (Grunfelder et al., 2003).

The positioning, replication and segregation of these organelles throughout the cell cycle is a tightly orchestrated process (Robinson et al., 1995). Analogous to the typical eukaryotic cell cycle model, trypanosomes pass through the G1, S, G2 and M phases. Perhaps predictably though, the parasites' cell cycle is distinguished by a number of novel control mechanisms and intricacies (Hammarton et al., 2007). The broad events of trypanosome cell cycle replication are described in Figure 1.4 (though most of the work on the trypanosome cell cycle has focused on the insect procyclic form of the life cycle, as shown in Figure 1.4, replicative bloodstream forms follow a similar sequence of events). The use of fluorescent cytological markers to stain the nuclear and kinetoplast DNA, such as the nucleic acid stain DAPI, allow for rapid and quantitative visualisation of cell cycle position via fluorescence microscopy or flow cytometry.

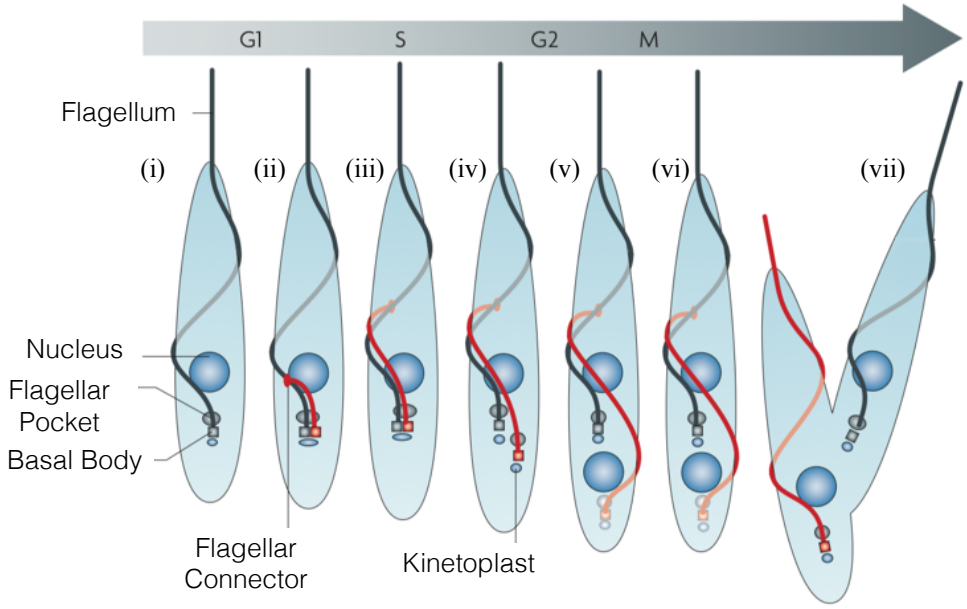


Figure 1.4 The cell cycle of procyclic form *T. brucei*. (i) In G1, cells have one kinetoplast and one nucleus (1K1N). The cell cycle begins with the elongation and maturation of the probasal body. The mature basal body is linked to the kinetoplast through a structure called the tripartite attachment complex, ensuring coordinated kinetoplast and basal body/flagellar replication. **(ii)** The basal body replicates in a microtubule mediated process, matures and a new flagellum begins to grow. In procyclic form trypanosomes, the new flagellum is connected via its tip to the axoneme of the original flagellum via the flagellar connector.

This connector mediates the transfer of structural information to the new flagellum via cytotaxic inheritance. In bloodstream forms there is no evidence of a flagellar connector. **(iii)** The growing daughter flagellum closely tracks the shape of the original flagellum. Kinetoplast (K) DNA replication commences just prior to that of nuclear (N) DNA replication and furthermore takes less time to complete. **(iv)** The basal bodies segregate, which in turn drives kinetoplast segregation. The cell is now 2K1N. **(v)** Nuclear mitosis occurs without breakdown of the nuclear envelope or chromosomal condensation. The cell becomes 2K2N. **(vi)** and **(vii)** Cytokinesis separates the cell into two 1K1N siblings. Cleavage furrow ingression initiates at the anterior pole of the cell and progresses along the longitudinal axis of the cell, passing through the two flagella. From (Field and Carrington, 2009; Hammarton et al., 2007; Matthews, 2005; Moreira-Leite et al., 2001; Robinson et al., 1995).

1.1.5 Trypanosome life cycle

Delivery of infectious metacyclic trypomastigotes to the dermal connective tissues by the bite of an infected tsetse fly initiates infection in the mammalian host (Figure 1.5). In the bloodstream, metacyclic trypomastigotes differentiate to proliferative long slender forms which rapidly increase in number. Differentiation is accompanied by repression of the mitochondrion as the parasites begin to generate ATP through the conversion of glucose to pyruvate (Haanstra et al., 2012). The trypanosome's glycolytic enzymes are housed within a unique membrane-enclosed organelle, the glycosome (Opperdoes and Borst, 1977). Proliferating cells evade the mammalian immune response via antigenic variation of their variant surface glycoprotein (VSG) coat (Horn, 2014)(see Section 1.2) and, upon reaching a critical cell density, differentiate to quiescent, transmission-competent short stumpy forms (Reuner et al., 1997) (see Section 1.3).

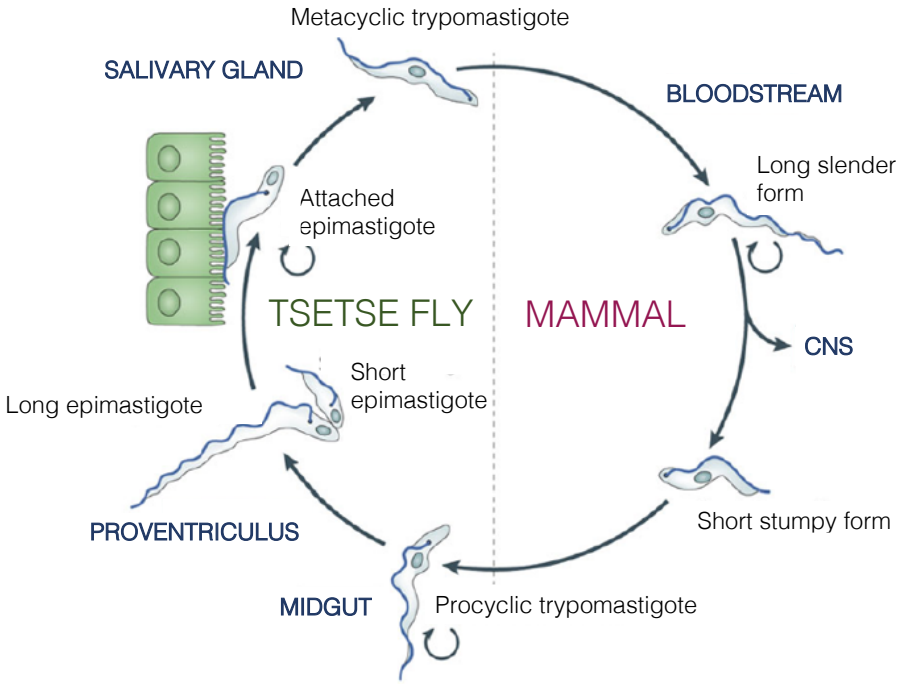


Figure 1.5 The major mammalian and tsetse developmental stages of the *T. brucei* life cycle. Circular arrows depict proliferative stages. Figure adapted from (Langousis and Hill, 2014).

Upon ingestion within a tsetse bloodmeal, stumpy cells differentiate synchronously to replicative procyclic forms within the endoperitrophic space of the midgut (Vickerman, 1965; Ziegelbauer et al., 1990). Over 48-72 hours, the VSG coat is progressively lost and replaced with a less dense, GPI-anchored GPEET and EP procyclin coat (Roditi et al., 1989; Turner et al., 1988b). GPEET expression is gradually decreased as the cells establish infection (Vassella et al., 2000). Glucose disappears from the bloodmeal within minutes, and as such procyclic cells display full mitochondrial activity after 14-24 hours as they begin to produce ATP through the oxidative phosphorylation of *L*-proline (Besteiro et al., 2005). Four days post bloodmeal, procyclic cells penetrate the peritrophic matrix and migrate to the proventriculus where they undergo gross morphological restructuring coupled to an asymmetric division, producing one short and one long epimastigote. Short epimastigotes attach to microvilli within the salivary gland mediated by dendritic outgrowths of their flagellum. Following extensive proliferation, epimastigotes

differentiate to metacyclic trypomastigotes and detach in preparation for delivery to the mammalian host. To aid immune evasion metacyclic trypomastigotes themselves acquire a variant surface coat. However, with at most 27 variants present, the available repertoire of metacyclic VSGs is far more limited than that of slender forms (Barry et al., 1983; Turner et al., 1988a).

In a significant discovery, it has recently been demonstrated that mammalian skin (Caljon et al., 2016; Capewell et al., 2016) and adipose tissues (Trindade et al., 2016) are additional anatomical reservoirs for bloodstream form *T. brucei*. Both infective replicating adipose resident parasites, termed adipose tissue forms (ATFs), and transmission-competent ATFs were found besides circulating bloodstream form parasites in mice. Though morphologically similar, the bloodstream and adipose resident trypanosomes were transcriptionally and metabolically distinct, with ATFs showing consistent upregulation of genes encoding enzymes involved in fatty acid β-oxidation (Trindade et al., 2016). Such an adaptation would allow for the ATFs to exploit a new environmental carbon source. The significance of these ATFs in the context of chronic sylvatic infection has yet to be confirmed, though they may contribute to the maintenance of tsetse transmission when blood parasitaemia is low, as is often seen in these reservoirs. Additionally, the nature of the tissue that they reside in might allow for more rapid accumulation of the density-dependent differentiation signal, stumpy induction factor (SIF, see Section 1.3) (Silvester et al., 2017).

1.2 Antigenic variation

Within the mammalian bloodstream *T. brucei* encounters a hostile environment, with the hosts' immune system constantly battling to control the infection. Yet despite this barrage of host innate and adaptive immunological defences, *T. brucei* has evolved to thrive. The parasite possesses the extraordinary capacity to periodically switch its expressed VSG coat to that of an antigenically distinct alternative, and thus evade antibody-mediated detection and clearance, via a process termed antigenic variation (Vickerman, 1978).

1.2.1 The VSG coat and host interactions

Over 10^7 glycosylphosphatidylinositol (GPI) -anchored VSG molecules of a single protein species form a densely-packed homogenous surface coat enshrouding the trypanosome. VSGs are homodimers of two 50-60kDa subunits, with each monomer consisting of a large exposed N-terminal domain (NTD) and a smaller concealed C-terminal domain (CTD) connected by a flexible linker, Linker 1 (Figure 1.6) (Carrington et al., 1991; Schwede et al., 2011). The NTD shows huge variation in its amino acid sequence, except for the localisation of a number of conserved cysteine residues, yet has a conserved tertiary structure (Carrington et al., 1991; Cross, 1996). In contrast, the CTD amino acid sequence remains relatively well conserved. Recent structural analysis has suggested that the linker region confers high flexibility, allowing the freely diffusing VSGs to adopt one of two different conformations, relaxed or compact. Such structural flexibility is hypothesised to allow the integrity of the protective barrier to be maintained in response to changes in protein density on the surface coat (Bartossek et al., 2017; Bulow et al., 1988).

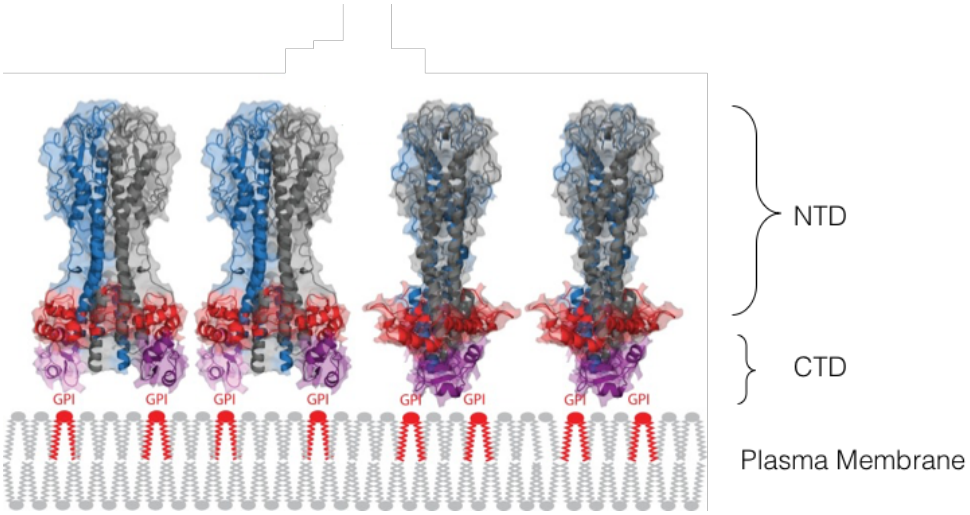


Figure 1.6 VSG structure. Each VSG monomer is highlighted in blue or grey, respectively, and the C-terminal domain in purple. The widest part of the VSG, for the left two VSGs this is ~5.9nm, whereas for the right two VSGs this is ~6nm, is highlighted in red. The VSG structure is derived from PDB 1VSG and 1XU6. Figure adapted from (Schwede et al., 2015).

The surface coat is the primary target of the immune response, and it is to epitopes within the highly immunogenic NTD that host antibodies are targeted (Hsia et al., 1996). Protruding ~12-15nm from the plasma membrane, VSGs serve to conceal small invariant surface antigens, such as transferrin and haptoglobin-haemoglobin receptors and hexose transporters, from antibody detection (Schwede et al., 2011). Modelling, however, has predicted that the coat is not completely impenetrable to host antibodies, with IgG predicted to reach within 5nm of the plasma membrane (Chattopadhyay et al., 2005; Mugnier et al., 2016; Schwede et al., 2011). Furthermore, two invariant surface proteins, ESAG 4 and ISG 65 were predicted to be similar in size, if not larger, than the VSGs themselves and thus likely accessible to antibodies (Schwede et al., 2011). Following a VSG switch, it has recently been estimated that parasites require ~4.5 days to completely replace their VSG coats and that for the initial 29 hours post VSG expression switch, the parasites remain vulnerable to host IgM antibodies (Pinger et al., 2017).

In addition to their shielding function, VSGs contribute to the removal of host antibodies from the cell surface. Bound immunoglobulins act as 'molecular sails' and are sorted by hydrodynamic flow to the posterior of the cell as the parasite swims forward (Engstler et al., 2007). From here the VSG-immunoglobulin complex is endocytosed in clathrin-coated vesicles at the flagellar pocket (Allen et al., 2003). Host antibody is degraded within endosomes and the VSG, now free of bound immunoglobulin, is trafficked back to the surface. At any one time it has been estimated that 90% of VSG proteins are present on the surface, whilst three quarters of the remaining 10% are present within endosomal compartments (Grunfelder et al., 2002). It has been suggested that parasite recognition and clearance is primarily mediated by a T-cell independent and IgM response. *In vivo* infections of $IgM^{-/-}$ mice have been shown to progress with similar kinetics as in wild type mice however, therefore implicating that immunoglobulins other than the IgM, specifically IgG, play a role (Black et al., 2010; Magez et al., 2008; Mugnier et al., 2016).

Despite continuous removal of surface-bound immunoglobulins, once a sufficient antibody titre is reached, the hydrodynamic flow is overwhelmed and the parasite cleared from the system. However, the population survives since VSG

switch variants which have arisen in the population, and express an antigenically distinct VSG, escape destruction and continue to proliferate. It is this periodic destruction of a given variant, followed by the outgrowth of antigenically distinct variants, which creates the characteristic undulating waves of parasitaemia typical of bloodstream trypanosome infection (Figure 1.7) (Horn, 2014).

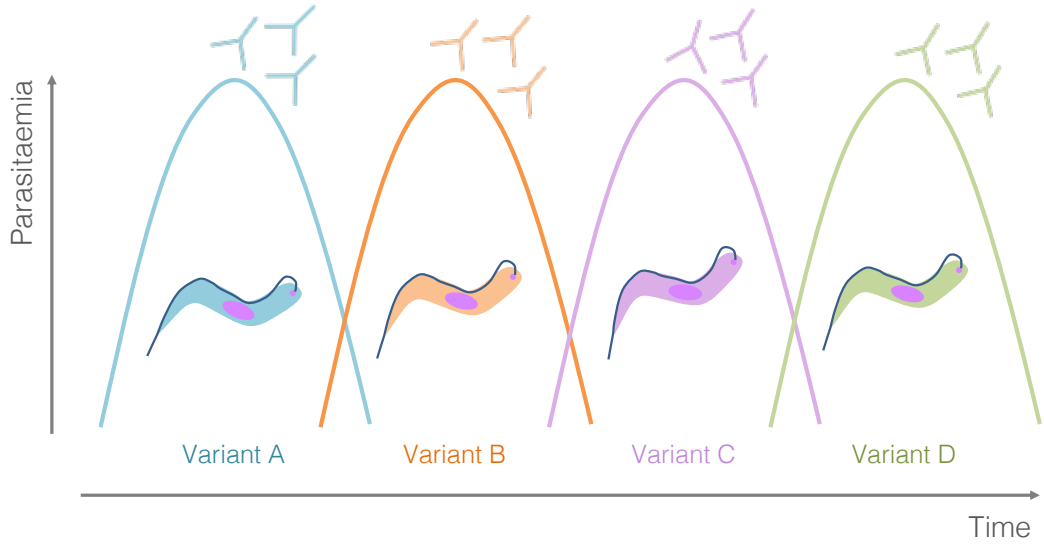


Figure 1.7 A simplified model of antigenic variation in trypanosomes. As the parasitaemia in the blood rises, immune responses are mounted towards specific variants which are ultimately cleared from the system. The infection continues, however, because antigenically distinct variants within the population escape detection and accumulate in number. In reality, each peak could be composed of ~100 different variants, though only half of these will surpass 1% of the total population (Hall et al., 2013; Mugnier et al., 2015).

1.2.2 The VSG archive

With antigenic variation playing such a pivotal role in the maintenance of chronic infection, it is unsurprising that trypanosomes have dedicated over 20% of their genome to the archive of VSG genes (Morrison et al., 2009). The trypanosome genome houses ~2000 different VSG genes, though only 5-10% of these actually encode intact and functional VSGs (Marcello and Barry, 2007). The rest are pseudogenes, which can only be expressed after recombining with another archival

VSG fragment to form a complete VSG gene, critical to the expansion of the potential VSG repertoire and maintenance of long-term infection (Hall et al., 2013). The majority of archived VSG genes are located in silent subtelomeric arrays on the 11 diploid megachromosomes, which are orientated away from the telomere and consist of 3 to ~250 (pseudo)genes (Barry et al., 2005; Berriman et al., 2005; Horn and Barry, 2005). Additionally, trypanosomes have evolved a highly specialised repertoire of ~100 30-150kb aneuploid mini-chromosomes and 1-7 larger 200-700kb intermediate-chromosomes, presumably to expand the VSG repository (El-Sayed et al., 2000; Wickstead et al., 2004). Mini-chromosomes contain a 20-80kb central core of tandem 177 bp repeats and most, though not all, house a single VSG gene at their subtelomere (Horn and Barry, 2005; Wickstead et al., 2004).

1.2.3 VSG expression sites

Expression of a VSG relies on the positioning of a VSG gene within one of ~20 telomere-proximal bloodstream expression sites (BES, Figure 1.8a) present on the megabase- and intermediate-chromosomes, of which only one is active at a time. Expression sites are large (~50kb) polycistronic transcription units, transcribed by RNA polymerase I (RNAPI) within a discrete extra-nucleolar structure known as the expression site body (ESB) (Navarro and Gull, 2001). Cotranscriptional *trans*-splicing produces the individual mRNAs (Huang and van der Ploeg, 1991).

The general framework of BESs is highly conserved (Berriman et al., 2002; Hertz-Fowler et al., 2008). Following an RNAPI promoter is up to 11 members of a diverse family of polymorphic genes and pseudogenes, termed expression site associated genes (ESAGs), the functions of very few of which are known (Berriman et al., 2005). ESAG 6 and 7 together encode a transferrin binding protein complex whilst ESAG 4 encodes a receptor-like adenylate cyclase which may be involved in regulation of the early immune response (Salmon et al., 2012; Steverding et al., 1994). Though the ordering of ESAGs on the BES largely remains the same (*ESAG7*, *6*, *5*, *3*, *4*, *8*, *3*, *2*, *11* and finally *ESAG1*), not all expression sites contain the complete repertoire of ESAGs, with gene duplications or deletions, pseudogenes and expression site truncations common occurrence (Hertz-Fowler et al., 2008; Rudenko,

2000). Such variation in ESAG expression between BESs is hypothesised to allow for species-specific adaptations. For example, it has been suggested that transferrin receptors encoded from different BESs have different transferrin binding affinities (Bitter et al., 1998; Gerrits et al., 2002; McCulloch et al., 2015), though this remains to be proved *in vivo* (Salmon et al., 2005).

Between the ESAGs and VSG gene, lies an extensive stretch of conserved AT-rich DNA repeats (so called '70 bp repeats', though in reality the repeats consist of a 76 bp consensus sequence) which act as a recombinogenic 'hotspot' (Campbell et al., 1984; Hovel-Miner et al., 2016). Approximately 1.5kb downstream from the repeats is the VSG gene, which itself is followed by ≥ 10kb of telomeric T_2AG_3 repeats (Horn and Barry, 2005).

A conserved telomeric region associated with VSG loci has recently been described in *T. congolense*, representing the first report of a VSG ES in that parasite (Abbas et al., 2018). The structures, present mostly on the minichromosomes but potentially also on the megabase chromosomes, were ~20-40,000bp in length and consisted of a conserved core 369bp repeat which stretched up to 25kb in length, conserved non-coding elements (CNEs), a VSG gene and, on ~20% of telomeric contigs, protein coding genes. These protein coding genes, however, were not orthologous to *T. brucei* ESAGs (Abbas et al., 2018). Though the telomeric region is not structurally homologous to the *T. brucei* BES, the subtelomeric positioning and functional similarities suggests that an ancestor common to *T. brucei* and *T. congolense* possessed a subtelomeric ES which has since diverged (Abbas et al., 2018)

1.2.4 VSG switching

Switching of the expressed VSG to an antigenically distinct alternative allows the trypanosome population to survive host immunological attack. An antigen switch may be facilitated by transcriptional changes amongst the BESs, or alternatively, through DNA recombination events (Figure 1.8b-e).

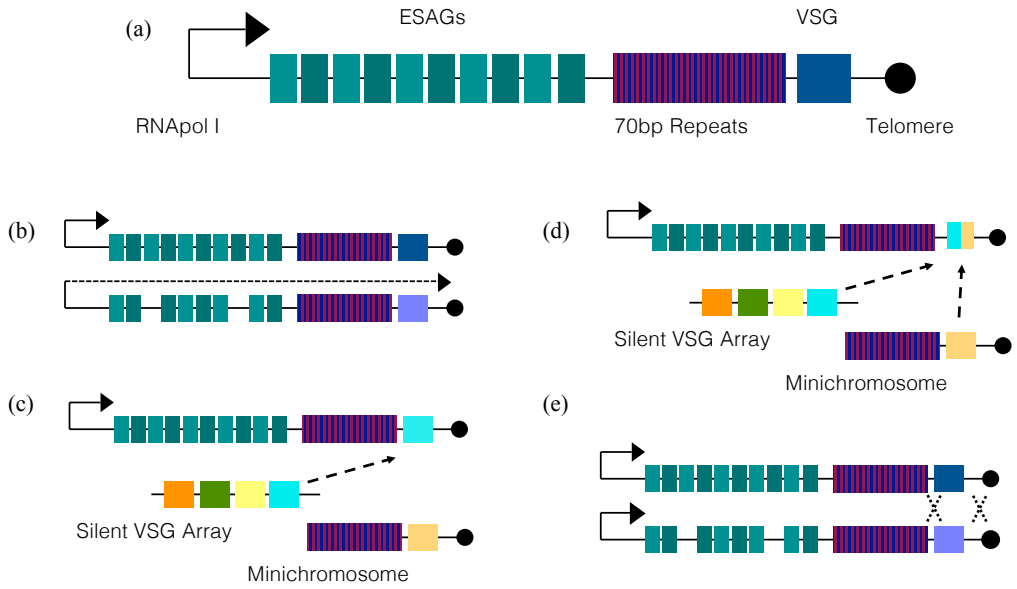


Figure 1.8 Mechanisms of antigenic variation within *T. brucei*. (a) A typical subtelomeric BES from the megabase chromosomes with the RNAP I promoter, ESAGS, 70 bp repeats, VSG gene and telomere depicted. **(b)** *In situ* transcriptional switching. **(c)** Duplicative gene conversion. **(d)** Segmental gene conversion. **(e)** Reciprocal telomeric exchange. Figure adapted from (Matthews et al., 2015).

Transcriptional switches, otherwise termed ES or *in situ* switches, encompass the silencing of an actively expressed BES and the subsequent activation of a single, previously silent, BES (Figure 1.8b) (Van der Ploeg et al., 1984). The strict state of monoallelic VSG gene expression following an *in situ* switch is controlled by a complex network of epigenetic control factors and is furthermore influenced by telomere-positional effects (reviewed in (Cestari and Stuart, 2018)). In addition to these epigenetic control factors, allelic control may be furthermore controlled at the transcription elongation and RNA processing level, since low levels of transcription still occur on silent BESs (Kassem et al., 2014).

Since the number of archival VSG genes is in vast excess to the number of BESs, it is perhaps unsurprising that the predominant means of VSG switching is expedited by DNA rearrangements (Robinson et al., 1999). These recombinatorial events exploit (a) sequence homology between VSG flanking regions (duplicative gene conversion) or coding regions (segmental gene conversion) and (b) the inherent

instability of the telomeric environment (reciprocal telomeric exchange) (Boothroyd et al., 2009; Horn, 2014; Pays et al., 1985; Pays et al., 1983).

Driven by homology within the 70 bp repeat regions flanking VSGs, duplicative gene conversion (GC) involves the duplication of an intact 'basic copy' silent VSG (from either another BES or from elsewhere within the genomic VSG archive) and the subsequent transposition of this expression linked copy (ELC) into the active BES (Figure 1.8c) (Campbell et al., 1984; Hovel-Miner et al., 2016). The displaced VSG is simultaneously deleted. Initiation of a GC VSG switch is still not fully understood, though induction of a DNA double-stranded break adjacent to the 70 bp repeat region was shown to increase switching *in vitro* ~250-fold and, furthermore, spontaneous DSBs could be detected within this region (Boothroyd et al., 2009). Why the 70 bp repeat region would be a specific target for DSBs has not yet been explained (McCulloch et al., 2015). The region upstream of the 70 bp repeats commonly delineates the 5' boundary of gene conversion, though Hertz-Fowler et al. reported in ~10% of VSG switches studied in monomorphic trypanosomes that the entire BES was replaced by duplicative transposition of a complete, new BES, including the promoter region (Campbell et al., 1984; Hertz-Fowler et al., 2008). Similar to, though seen less frequently than GC, is reciprocal telomeric exchange (Figure 1.8e). Here, the telomere of a silent BES is exchanged via cross-over for that of the active BES, though unlike in GC the VSG gene is not deleted and instead remains on the silenced telomere (Pays et al., 1985). Imperative to the expansion of the VSG repertoire is segmental gene conversion, otherwise known as 'mosaic formation'. Mosaic VSGs are assembled upon the recombination of at least two archival VSG gene or pseudogene ORFs to create a novel coding VSG gene (Figure 1.8d). The accumulation of segmental gene conversion events generates strings of VSG variants that exhibit a degree of sequence homology, yet are antigenically distinct (Hall et al., 2013; Lythgoe et al., 2007). Despite building evidence that mosaic VSG formation is critical to maintenance of long-term infection (discussed in Section 1.2.5), our understanding of the substrates and mechanisms of mosaic VSG formation are currently limited. Where mosaic assembly occurs and what the substrates for recombination are, are both important questions that remain unanswered.

VSG expression is semi-predictable and hierarchical, as demonstrated by the ordered, preferential expression of certain subsets of VSG genes within the first relapse peaks of parasitaemia, and across the time course of a chronic infection (Morrison et al., 2005). Whilst the host immune response does select for the expression of novel VSG variants, the hierarchy of VSG expression is a parasite intrinsic process that is not driven by the host. The probability that a given VSG gene will be activated is highly influenced by the locus type that it resides in and the mode of switching (Gjini et al., 2010; Morrison et al., 2005). *In situ* transcriptional switches are favoured early in infection, followed by gene conversion of archived intact silent genes into the active BES and finally, mosaic VSG generation via segmental gene conversion dominates late infection. More recently, the length of the expressed VSG has been shown to be fundamental to the generation of an expression hierarchy (Liu et al., 2018). Mathematical modelling predicted parasites expressing VSGs of a short length to accumulate more quickly early in infection compared to parasites expressing longer VSGs, hypothetically due to the lower metabolic burden of producing shorter VSGs. Surpassing the threshold for an immune response first, the cells expressing shorter VSGs would act as a 'distraction' allowing the slower growing cells expressing longer VSGs to accumulate in the absence of immune attack. Such predictions were backed up by *in vivo* and RNAseq data, whereby it was found that shorter VSGs accumulated early in infection, whilst longer VSGs predominated during chronic infection in the mouse model (Liu et al., 2018).

1.2.5 Regulation of VSG expression

RNAPI transcription initiates simultaneously on multiple BESs, progressing at least as far as ESAG6 before transcription is terminated in all but the active VSG ES, resulting in the expression of only one VSG surface coat (Kassem et al., 2014; Vanhamme et al., 2000). The control and regulation of monoallelic VSG expression is a complex and multifactorial process which incorporates a number of epigenetic mechanisms and is influenced by the nuclear architecture and a developmentally-regulated telomere position effect (Cestari and Stuart, 2018; Horn and Cross, 1995).

The eukaryotic nucleus is organised into transcriptionally 'silent' regions of condensed heterochromatin and transcriptionally 'active' regions of open euchromatin. Indeed, it has been observed in *T. brucei* that electron dense heterochromatin clusters at the nuclear periphery whilst the electron lucent euchromatin is located in the nuclear interior (Ogbadoyi et al., 2000). Inactive telomeric BESs are found in this transcriptionally repressed heterochromatic region at the nuclear periphery, whilst the single active VSG BES is transcribed by RNApol I at a distinct, transcriptionally permissive, extranucleolar site termed the ESB (Navarro and Gull, 2001). Metazoan nuclear lamins are critic for maintaining the structural integrity of the nucleus, nuclear pore complex (NPC) positioning and organisation of the genome into heterochromatin and euchromatin. In *T. brucei*, the 450kDa coiled-coil protein Nuclear Peripheral Protein 1 (NUP-1) is an essential component of fibres at the inner nuclear envelope and is functionally analogous to metazoan lamins (DuBois et al., 2012). Knock-down of NUP-1 resulted in the transcription of previously silent telomere proximal BESs and procyclin loci and a 10-fold increase in VSG switch rate, demonstrating that the structural protein plays a role in silencing of telomeric genes at the nuclear periphery and the control of antigenic variation (DuBois et al., 2012). A functional interacting partner of NUP-1, NUP-2, was also found to play a role in the silencing of telomeric BESs, however its knock-down did not cause an increase in VSG switching (Maishman et al., 2016).

Chromatin remodelling is integral to monoallelic VSG expression and as such, the chromatin structures of active and silent BESs are distinct. Active BESs are depleted in nucleosomes, indicating an open chromatin state which facilitates transcription (Figueiredo and Cross, 2010; Stanne and Rudenko, 2010). In contrast, the histones H2AZ, H3 and H4 are enriched on the silent BESs and non-transcribed 50bp and 177bp repeat arrays, indicating that the chromatin is in a condensed, closed confirmation and therefore does not facilitate transcription (Figueiredo and Cross, 2010; Lowell et al., 2005). A number of histone chaperones, which regulate the assembly of nucleosomes, have been implicated in the control of BES transcription. The 'facilitates chromatin transcription' (FACT) chromatin remodelling complex, in *T. brucei*, is composed of two subunits: Spt16 and Pob3 (Denninger et al., 2010; Denninger and Rudenko, 2014). Chromatin immunoprecipitation (ChIP)

demonstrated that Spt16 was enriched at the promoters of transcriptionally silent BESs. Blocking the synthesis of the Spt16 subunit opened the chromatin structure around these BES promoters and thus caused a 20-23-fold derepression of silent BESs (Denninger et al., 2010; Denninger and Rudenko, 2014). This derepression was accompanied by an accumulation of cells arrested in the G2/M stage of the cell cycle (Denninger et al., 2010), however by blocking cyclin 6 synthesis to create a mitotic block (Hammarton et al., 2003) and observing that silent BESs were not derepressed, Denninger and Rudenko (2014) demonstrated that VSG ES derepression was not a general feature of G2/M arrest. Knock down of Spt16's partner Pob3, also caused the derepression of VSG BESs (Denninger and Rudenko, 2014). To identify candidate chromatin regulators which might act as BES regulators, Alsford and Horn (2012) analysed their previously generated RITseq data (Alsford et al., 2011) for genes which displayed a high loss of fitness phenotype when knocked down in BSF cells. The histone chaperones anti-silencing factor 1A (ASF1A) and chromatin assembly factor 1B (CAF-1b) were identified in this screen (Alsford and Horn, 2012). ASF1A and CAF-1b were knocked down by RNAi in VSG cell lines which contained a promoter-proximal GFP ORF in a repressed BES. Following 24 hours of induction, there was an increase in GFP expression, demonstrating that derepression of the silent VSG ESs had taken place. Whilst derepression was evident at a promoter proximal site, there was no evidence of derepression at the promoter-distal sites, such as the VSG gene. This shows that factors downstream of ASF1A and CAF-1b must also be involved in ES transcriptional control (Alsford and Horn, 2012; Cestari and Stuart, 2018).

Post-transcriptional modifications of histones by methylation, ubiquitylation, phosphorylation or SUMOylation alter chromatin structure and hence, gene expression. Disruptor of telomeric silencing 1B (DOT1B) is a histone methyltransferase which trimethylates lysine 76 of histone H3 (Figueiredo et al., 2008). Deletion of DOT1B causes the partial derepression of silent telomeric BES sites without affecting non-telomeric loci whilst furthermore permitting the expression of two VSG surface coats simultaneously (Figueiredo et al., 2008), thus demonstrating that DOT1B is required for monoallelic expression. The kinetoplastid-specific modified base β-D-glucosyl-hydroxymethyluracil, otherwise known as Base

J, is a chromatin mark present in BSF, but absent from PCF trypanosomes (van Leeuwen et al., 1997). The base is synthesised in two steps: thymidine in the DNA is firstly hydroxylated by two thymidine hydroxylase enzymes, JBP1 and JBP2 (Cross et al., 2002; DiPaolo et al., 2005), before the glucosyltransferase JGT catalyses the transfer of a glucose moiety from uridine diphosphoglucose to the 5-hydroxymethyluracil intermediate (Bullard et al., 2014). The presence of Base J correlates with the silencing of BESs (van Leeuwen et al., 1997). Base J was found to be abundant in the 50bp repeats upstream of the BES promoter and the telomeric repeats downstream of the VSG gene of both active and silent BES, however in silent BESs the modification was found to be present along the entire length of the BES too, including within the 70bp repeats and VSG gene itself (van Leeuwen et al., 1997).

In an impressive recent study, Muller et al. (2018) demonstrated a link between genome architecture and antigenic variation mediated through the histone variants H3.V and H4.V. Single cell RNAseq (scRNA-seq) and Pacbio single-molecule real-time (SMRT) sequencing showed that deletion of both variants (ΔH3.VΔH4.V) induced a VSG switch in 74% of analysed cells through the recombination of a pair of ESAG 8 genes present on BES1 and BES15 (Muller et al., 2018). Deletion of histone variant H3.V or H4.V alone did not induce a VSG switch. The mechanism behind this increase in VSG switching by DNA recombination was investigated using genome-wide chromosome confirmation capture (Hi-C) and assays for transposase-accessible chromatin (ATAC-seq). Deletion of histone variant H3.V alone grossly changed the genome architecture and furthermore promoted increased interactions among the silent BESs. Whilst deletion of histone variant H4.V did not induce significant changes in the genome architecture, there was a marked increase in promoter-proximal DNA accessibility upon its deletion (Muller et al., 2018). Therefore, the histone variants H3.V and H4.V are involved in regulating DNA accessibility and genome architecture and deletion of both induces changes in VSG expression (Muller et al., 2018).

VSG exclusion factor 1 (*VEX1*) was identified in an RNAi library screen for *T. brucei* genes that control allelic exclusion (Glover et al., 2016). VEX1 was

contained within an extranucleolar compartment which, in BSF cells, was closely associated with the RNAPI focus of the active BES. In PCF cells, however, VEX1 was redistributed and had a broad nuclear staining pattern (Glover et al., 2016). Upon either RNAi knockdown or ectopic overexpression of VEX1 in BSF cells, there was an accumulation of cells expressing multiple VSGs on their surface, as detected by quantitative proteomics. RNAseq analysis of the expressed transcriptome in *VEX1* RNAi cells found that there was >26-fold increase in transcripts encoding ES-associated VSGs, in addition to an increase in transcripts encoding metacyclic VSGs (which are also transcribed by RNAPI) and genes immediately adjacent to VSGs (Glover et al., 2016). The transcriptomes of VEX1 overexpressors were also enriched >21-fold for ES-associated VSGs, however there was also a marked increase in transcripts encoding ESAGS (from both the active and silent ESs) and non-telomeric RNAPI transcribed genes such as *procyclin*. This therefore shows that VEX1 negatively regulates VSG genes, and positively regulates VSG genes and non-telomeric RNAPI transcribed genes (Glover et al., 2016). A 'winner-takes-all' mechanism of monoallelic control was proposed by the authors, whereby VEX1 sequestration at the active ES enables RNAPI transcription of the entire BES whilst simultaneously repressing transcription of the other ESs. Repression of the ESs was hypothesised to occur in a VEX1-dependent form of homology-dependent silencing (Glover et al., 2016).

Although the mechanisms controlling monoallelic VSG expression primarily occur at the level of transcription elongation, a number of factors can influence the initiation of RNAPI transcription too, such as class I basal transcription initiation factor A (CIFTA) and the presence of the posttranslational modification, small ubiquitin-like modifier (SUMO) (Cestari and Stuart, 2018; Lopez-Farfan et al., 2014; Nguyen et al., 2014). For a more comprehensive review of the genes implicated in the control and regulation of VSG expression, see Cestari and Stuart (2018).

1.2.6 VSG diversity within chronic infection

The kinetics of VSG switching, and thus immune evasion, predictably impact the dynamics of mammalian infection, though until recently the full extent to which

these exert their influence had been poorly understood. The recent application of high-resolution, quantitative sequencing approaches to longitudinal VSG switch studies has helped to shed light on VSG switching dynamics and diversity within chronic infection.

In theory, a trypanosome should switch its expressed VSG at a rate quick enough to avoid immune detection and clearance, but not so fast as to rapidly deplete their VSG repertoire of ~400 intact genes. Recent evidence however, has suggested a far greater level of variant diversity within individual infections, including within the parasitaemic valleys, than was previously appreciated, this being in excess of what would be predicted to 'only just' avoid immune detection (Mugnier et al., 2016). Hall et al. (2013), using a cloning and RT-PCR approach, measured ~20 variants/peak but predicted as many as 100 variants could be expressed in a single parasitaemia peak (Hall et al., 2013). Similarly, a targeted RNAseq approach, termed VSGseq, adopted by Mugnier et al. (2015), averaged 28 variants/peak during the first 30 days of infection, with a maximum of 83 variants within a single peak being sequenced (Mugnier et al., 2015). Each peak consisted of a handful of 'major' VSG variants which made up the majority of the population and a number of rarer, low abundance 'minor' VSG variants, about half of which would never surpass 1% of the population (Mugnier et al., 2015). In accordance with the aforementioned hierarchy of VSG gene expression, and despite extreme levels of VSG expression diversity, the same VSG variants frequently occurred during different infections (Hall et al., 2013; Mugnier et al., 2015). However, of the 48 VSG genes that occurred in every infection in the Mugnier study, there was rarely a gene which was always dominant, and conversely nor were there any VSG genes which were consistently rare (Mugnier et al., 2015). Infections in natural hosts can last from months to years, and with a new variant peak consisting of at least 20 variants arising every 5-8 days it is not difficult to appreciate that the number of expressed VSGs is obviously far in excess of the number of the ~400 intact coding VSG genes. Thus, following exhaustion of this intact repertoire, segmental gene conversion comes to dominate (Hall et al., 2013). Whilst these studies are undoubtedly informative, it must be noted that the mouse model, rather than a natural livestock host was employed for these experiments. How well these models relate to natural livestock infections, where *T.*

vivax and *T. congolense* predominate, remains to be seen (McCulloch et al., 2017) particularly because the emphasis of mosaic-gene conversion seems much less in these species. In support of the models however, a study of plasma pooled from three Ugandan patients infected with *T. b. rhodesiense* found a surprisingly diverse range of VSGs to be the most predominantly identified proteins (Eyford et al., 2013; Mugnier et al., 2015).

The interaction, if any, between antigenic variation and the slender to stumpy differentiation capacity within a chronic infection also remains to be fully understood. Infections are dominated by cell cycle arrested, transmission competent stumpy cells which do not switch their expressed VSG (MacGregor et al., 2011). How such a high level of VSG diversity is maintained during chronic infection despite the predominance of these cells, which were originally thought to limit exposure of the full antigen repertoire, poses an interesting question. It has been proposed that stumpy cells may contribute to the persistence of rare variants by limiting parasite numbers to a level below which the host immune system will mount a response (Hall et al., 2013).

1.2.7 Measuring the rate of antigenic variation

Distinct experimental trends have shaped the study of antigen switch frequency in African trypanosomes. Beginning with observations of small numbers of switch variants arising in chronic mammalian BSF infections using limited panels of VSG-specific antisera, series' of experiments have subsequently adopted ever more sophisticated (and less laborious) means of detecting VSG switch variants.

Early VSG switch frequency investigations traditionally studied parasites isolated from chronic *in vivo* infections (Gray, 1965; Myler et al., 1985) (though in order to demonstrate that antigenic variation occurred in the absence of an immune system, Doyle et al. studied parasites for up to 60 days of *in vitro* culture (Doyle et al., 1980)). Such experiments routinely isolated parasites from the animal (on a daily or weekly basis) and screened ~1000 cells for the expressed VSG by IFA using small libraries of VSG-specific antisera. Assessing VSG switch rate from chronic

infections is not only laborious, but complicated by the fact that it cannot be ascertained if a given variant arose from the variant preceding it, or if it is simply the progeny of a minor VSG variant already present, but previously undetectable, in the infection (Aitcheson et al., 2005). Single relapse experiments in rodents, in which a single switch away from the starting VSG is monitored, resolve this uncertainty. Miller and Turner (1981) performed the classic single relapse experiment, whereby 10 homogenous trypanosomes (with regard to their VSG expression) were injected into naïve rats. Parasitaemia was monitored until a second (relapse) peak of parasitaemia occurred, the mice exsanguinated and the parasites analysed for VSG expression based on immunofluorescence staining with VSG-specific anti-sera (Miller and Turner, 1981). Using this method, Miller and Turner analysed 47 first relapse populations of six clones and found that certain VSG types were expressed more frequently in the relapse peak than other, rarer, variants (preceding the finding that VSG expression is hierarchical). Later studies performed a variation on these experiments, such that immune selection against the starting VSG was mediated either by prior immunisation of the experimental rodent (Hartley and McCulloch, 2008; McCulloch and Barry, 1999; McCulloch et al., 1997; Proudfoot and McCulloch, 2005; Robinson et al., 2002) or by incubation of the parasites with serum and a VSG-specific antibody prior to injection (Lamont et al., 1986; Turner, 1997). If a rodent developed a patent parasitaemia, this indicated that a VSG switch variant had been present in the initial inoculum. Perhaps the biggest limitation of experiments exploiting the immune system to clear non-switched VAT variants, in addition to the need of prior immunisation of animals, is the potential for off-target responses. Thus, if the experimental animal is not immune-suppressed, the innate immune system may target cells for clearance in a non-specific manner and furthermore, antibodies (whether in the rodent or *in vitro* neutralisation assays) could bring about the complement-mediated lysis of all cells expressing similar epitopes to the starting VSG, thereby potentially causing some variants to be lost.

Many methods have been described to quantitate VSG switching *in vitro*, thus removing the need to work with large numbers of immunised animals and furthermore isolating the parasite from host factors which may complicate the interpretation of assays. Since the laboratory adapted strains of parasites used in

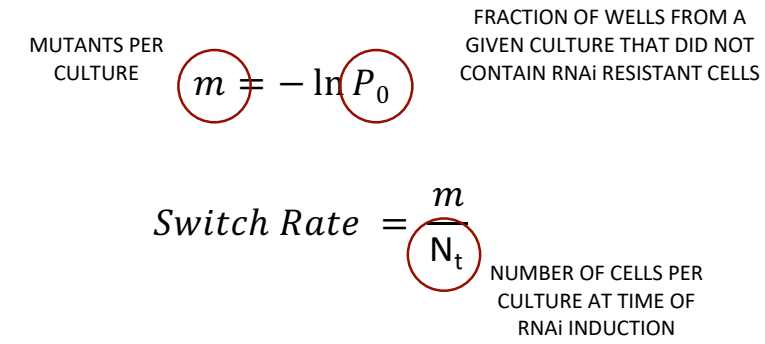
these studies switch their expressed surface VSG at relatively low rates, selection for VSG switch variants, or depletion of non-switched cells from the population is required to aid detection. One such strategy, at its simplest, involves marking both the active and a second ES with different drug resistance genes (Figueiredo et al., 2008; Hertz-Fowler et al., 2008; Horn and Cross, 1997). Following a period of growth in non-selective medium, the populations are treated with the drug whose resistance is encoded in the second ES, and are then cloned by limiting dilution. This positive drug-selection isolates resistant cells which have stochastically switched their formerly expressed VSG (for e.g. VSG 118) to that of the second VSG (for e.g. VSG221) via an *in situ* switch. Such a strategy is obviously limited as it only permits the study of the activation and inactivation of two marked ESs. A second, and far more thorough, experimental strategy for measuring VSG switch rate was described by Aitcheson et al. (2005). RNAi-mediated knockdown of the actively expressed VSG causes rapid precytokenesis arrest in trypanosomes *in vitro* (Sheader et al., 2005) but does not induce as antigenic switch, as confirmed by fluctuation analysis (Luria and Delbruck, 1943; Sheader et al., 2005). Thus, Aitcheson et al. (2005) adopted an RNAi strategy for their *in vitro* switch assays in order to select for cells which had switched away from the major VSG221 variant in the absence of immune selection. This method was advantageous since it rapidly generated high numbers of clonal switch variants, was not restricted to monitoring switching between two ESs only and depleted the starting variant with high specificity, though it did rely on tight regulation of the RNAi machinery. Briefly, serially diluted cultures of VSG221 expressing trypanosomes were expanded for 3 generations in 96-well plates to allow for an antigenic switch to occur before RNAi against the starting VSG was induced. Non-switched trypanosomes rapidly arrested, whilst those that had switched, either by DNA rearrangements or an *in situ* switch continued to proliferate and could be characterised following 8-10 days of further culture (Aitcheson et al., 2005). Immunofluorescence and PCR amplification of marker genes within the ESs facilitated the identification of the VSG switch mechanisms utilised by 127 RNAi resistant clones. Similarly, in cell lines transfected with a copy of the Herpes Simplex Virus thymidine kinase (HSV-TK) gene in the region between the 70bp repeats and the actively expressed VSG, treatment of serially diluted cultures with the drug

Ganciclovir can also select for cells which have switched their expressed VSG, and therefore do not express HSV-TK, during the course of an assay (Devlin et al., 2016; Kim and Cross, 2010). Since 2009, depletion of trypanosomes from the population which have not switched their VSG during an experiment has increasingly been performed by magnetic-activated cell sorting (MACS) (Boothroyd et al., 2009; Hovel-Miner et al., 2016; Hovel-Miner et al., 2012; Kim and Cross, 2010; Schulz et al., 2016). Following a period of *in vitro* growth in non-selective conditions, live cells are incubated with magnetic beads coupled to an antibody recognising the starting VSG (Boothroyd et al., 2009; Schulz et al., 2016). The cells are passed through a column to which the coated non-switched trypanosomes bind, and VSG switched cells pass through. The switched cells are collected in the flow-through and cloned by limiting dilution in 96 well plates. Though MACS separation does represent a quick alternative to selection with antibiotics or antibodies, it has been acknowledged that the many centrifugation steps can lead to shedding of the surface coat VSG (Schulz et al., 2016). Another complicating factor is that all steps must be performed at 4 degrees Celsius in order to circumvent internalisation of the antibody to the starting VSG. To confirm that MACS-enrichment for VSG switched cells did not bias results, Kim and Cross (2010) performed their VSG switch assays both with and without MACS-enrichment for VSG switched cells and found that the fold increase in VSG switching in the absence of TOPO3α was the same whether MACS was (10 to 40-fold) or was not (10 to 30-fold) used (Kim and Cross, 2010).

Just as the experimental methods vary between studies, so too does the method used to calculate VSG switch frequency. The majority of recent VSG switch studies have employed a simple approach, whereby the switch rate is presented as the number of VSG switched cells divided by the total number of cells (which can then be normalised to the number of parasite generations elapsed during the assay to obtain the VSG switches/cell/generation) (Boothroyd et al., 2009; Devlin et al., 2016; Hertz-Fowler et al., 2008; Horn and Cross, 1997; Hovel-Miner et al., 2016; Hovel-Miner et al., 2012; Kim and Cross, 2010; Schulz et al., 2016). This approach, however, is perhaps best suited to measuring relative change in VSG switch frequency following genetic perturbations, rather than absolute switch frequency,

since a number of variables are not accounted for. The calculation simply expresses the proportion of the population which was not expressing the starting VSG at the time of sampling. What is not accounted for is the phenotypic or genetic diversity within those VSG switched cells (Hovel-Miner et al., 2012). A 'negative' cell could have switched expression of its surface VSG twice since expressing the starting VSG or two cells which appear phenotypically similar could have switched by different means to get to the same VSG. This method would not account for either of those differences. Finally, there is some experimental evidence to suggest that different VAT variants may grow at different rates to one another (Lamont et al., 1986; Liu et al., 2018; Miller and Turner, 1981). By normalising the calculated VSG switches/cell to the total number of generations measured for the population as a whole, this may not account for these differences in growth between VAT variants. A calculation which does account for different growth rates (providing clones do not switch at a high rate) is the P_0 method, developed by Luria and Delbruck (1943) and used to calculate VSG switching frequency by Lamont et al. (1986), Turner (1997) and Aitcheson et al. (2005) (Figure 1.9). This method carries the requirement *in vitro* that non-switched cells must efficiently be depleted from the population (whether by immunological means, RNAi or MACS) and that cultures are initiated from a single VAT variant.

IN VITRO, AITCHESON ET AL. (2015)



IN VIVO, TURNER (1997)

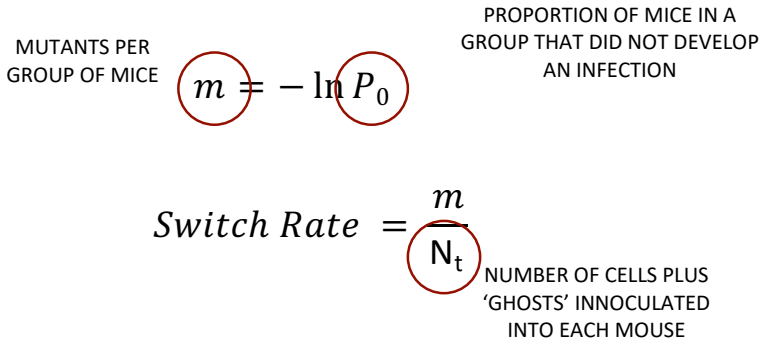


Figure 1.9 The P_0 method of calculating VSG switch frequency From Turner (1997) and Aitcheson et al. (2015). For Turner experimental method see Section 1.4. Based on fluctuation analysis methods established by Luria & Delbruck (1943).

In the majority of the above *in vitro* experiments, marker genes (primarily) in the starting VSG ES have aided characterisation of the switched clones. IFA analysis of the expressed VSG or a promoter-proximal GFP (Aitcheson et al., 2005; Figueiredo et al., 2008) can confirm that the selection for VSG switchers was successful and report on ES activity, respectively. To establish whether the clones switched by GC or an *in situ* switch, clones can be tested in drug resistance assays (Hovel-Miner et al., 2016; Hovel-Miner et al., 2012; Kim and Cross, 2010) or their DNA analysed by Southern Blot or PCR for the presence of the original expression-linked copy of the VSG (Aitcheson et al., 2005; Boothroyd et al., 2009; Devlin et al., 2016; Hovel-Miner et al., 2016; Hovel-Miner et al., 2012; Kim and Cross, 2010).

Other studies have defined switch mechanism using RAGE (Boothroyd et al., 2009) or Western blot analysis of marker expression (Devlin et al., 2016). A disadvantage to methods such as IFA or PCR is that their labour and time intensive manner dictates that limited numbers of clones arising in switch assays can be characterised. In general, assays characterise tens to hundreds of cells by IFA (though Turner & Barry (1989) claim between 24,000 and 153,000 cells/group of 6 mice were counted by IFA in their experiments). Flow-cytometry detection of VSG expression can circumvent this constraint. Tens of thousands of cells can be monitored for VSG expression in less than a minute and, in addition to processing superior sample sizes, flow-cytometry also has the advantage of being able to quantitate multiple markers simultaneously. Therefore, with the correctly placed markers, a study could not only just quantitate VSG switch rate but furthermore determine, for example, if a specific genetic perturbation increased the rate of DNA recombination based switches. Boothroyd et al. (2009), Hovel-Miner et al. (2012, 2016) and Schulz et al. (2016) have all reported on VSG switch rates following FACS analysis of VSG expression following MACS.

The low antigen switch frequency of many laboratory-adapted strains of trypanosomes (~1 switch in every 100,000 to 10,000,000 cells) can make achieving sufficient depth of analysis in VSG switch assays difficult. As such, many studies have elected to artificially increase the rate of VSG switching in cell lines by introducing an inducible double strand break (DSB) site into the actively expressed VSG ES (Boothroyd et al., 2009; Glover et al., 2013a). This is particularly desirable when large numbers of clones are required for characterisation. A polymorphism within the VSG 221 ES allows the specific integration of an *I-Sce*I restriction site directly adjacent to the 70bp repeats preceding the VSG gene, whilst a copy of the I-*Sce*I meganuclease ORF under the control of an inducible promoter is integrated at a silent site within the genome (Glover et al., 2013a). Addition of tetracycline induces the expression of I-*Sce*I, which forms a DSB at the restriction site within the active VSG ES. This break is repaired through 70bp repeat recombination, which translocates a new VSG gene into the ES, therefore mediating a GC switch. Transcription of the ES is not altered. Upon induction of I-*Sce*I expression, Boothroyd et al. (2009) predicted that ~1% of cells underwent a GC VSG switch,

with the VSG switch frequency rising ~250 fold to $1.5x10^{-3}$ compared to an uninduced control ($1.5x10^{-5}$). The donor VSGs which were translocated into the active ES predominantly resided in another ES, though 3 of the donor VSGs (~16%) had arisen from the minichromosomes (Boothroyd et al., 2009). PCR analysis of the region between VSG221 and the ES telomere found that, in I-*Sce*I-mediated switched clones, a ~500bp subtelomeric region which was present in unswitched clones was lost, therefore suggesting that resolution of the DSB was mediated by break induced replication (BIR) rather than crossover events in the 70bp repeats and 3' end of VSG 221 (Boothroyd et al., 2009). Induction of I-*Sce*I expression in the absence of the 70bp did not result in an increase in VSG switching, thus suggesting that the 70bp repeats facilitated repair of the DSB through homologous recombination (Boothroyd et al., 2009). Making use of this cell line, and a series of 70bp mutants, Hovel-Miner et al. (2016) were able to confirm Boothroyd's hypothesis that homologous recombination through the 70bp repeats promoted access to the VSG archive (Hovel-Miner et al., 2016). Using a slightly different I-*Sce*I expression system and VSG 221 ES construct, Glover et al. (2013) instead found that 100% of induced cells underwent a VSG switch when the I-*Sce*I RS was adjacent to the 70b repeats upstream of VSG 221. When the I-*Sce*I RS was situated adjacent to either the ES promoter or the telomeric T_2AG_3 repeats, VSG switching was far less efficiently induced (8% and 28% of clones switched their VSG upon induction of I-*Sce*I expression, respectively) (Glover et al., 2013a). The position of the DSB within the ES not only influenced the probability of a switch but furthermore the mechanism by which a VSG switch would be mediated. As Boothroyd et al. (2009) had previously found, a DSB break adjacent to the 70bp repeats upstream of the VSG gene resulted in the majority of VSG switches being gene conversions facilitated through 70bp homologous recombination. In contrast, when the DSB was induced adjacent to the ES promoter or telomere, DNA recombination rarely occurred and instead most clones either lost or replaced the ES entirely (Glover et al., 2013a).

Moving forward, it seems that the next experimental phase of VSG switching studies will involve the use of deep sequencing methods in order to determine expressed VSG diversity within chronic infections, identify the mechanism(s) of

segmental gene conversion, investigate VSG expression diversity within specific sub-compartments of the mammalian host, and characterise VSG switching and expression diversity in *T. congolense* and *T. vivax* (McCulloch et al., 2017). Such work would build on previous studies by Hall et al. (2013) and Mugnier et al. (2015) who both demonstrated the VSG expression diversity during chronic infection is far greater than previously imagined, and that SGC is the mediator of this expression diversity using the methods described in Section 1.2.6.

1.3 Slender to stumpy differentiation

Whilst the dynamics of the hosts' immune response and the trypanosomes' capacity for antigenic variation undoubtedly act together to influence the profile of the cyclical waves of parasitaemia, a second biological event within each of these waves, the density-dependent differentiation of parasites, also contributes strongly (Gjini et al., 2010). An important evolutionary adaptation, this developmental switch both restricts parasitaemia, thus prolonging infection, and preadapts the parasite for survival upon uptake by the tsetse fly vector.

1.3.1 The slender to stumpy transition and transmission potential

Upon injection by the tsetse fly, infections are established in the host bloodstream by rapidly proliferating morphologically slender forms, which quickly increase in number. Allowed to replicate unchecked, slender parasites would rapidly kill a host, thus limiting the parasite's transmission potential (MacGregor et al., 2011). As a consequence, it is essential that parasites regulate their growth *in vivo* to promote longevity of the host and maximise transmission potential, therefore allowing the successful continuation of the parasite's lifecycle by tsetse fly uptake (MacGregor et al., 2011). Such regulation is achieved by a developmental switch upon reaching a critical parasite density. During this developmental switch, proliferative slender forms respond to the accumulation of a small soluble parasite-derived factor (or factors) termed stumpy induction factor (SIF) and differentiate to specialised cell-cycle arrested stumpy forms (Hamm et al., 1990; Reuner et al., 1997;

Vassella et al., 1997a). Stumpy cells are irreversibly and uniformly arrested in G1/G0 (Shapiro et al., 1984), and since their abundance increases in response to ascending parasitaemia, the parasite population is restricted - prolonging the infection and optimising the transmission of the disease (MacGregor et al., 2011). The rapid increase in parasitaemia facilitated by slender form proliferation, commitment to cell-cycle exit, morphological transformation to stumpy cells and the subsequent clearance of quiescent cells by the immune system therefore contributes to the typical undulating trypanosome profile (Figure 1.10).

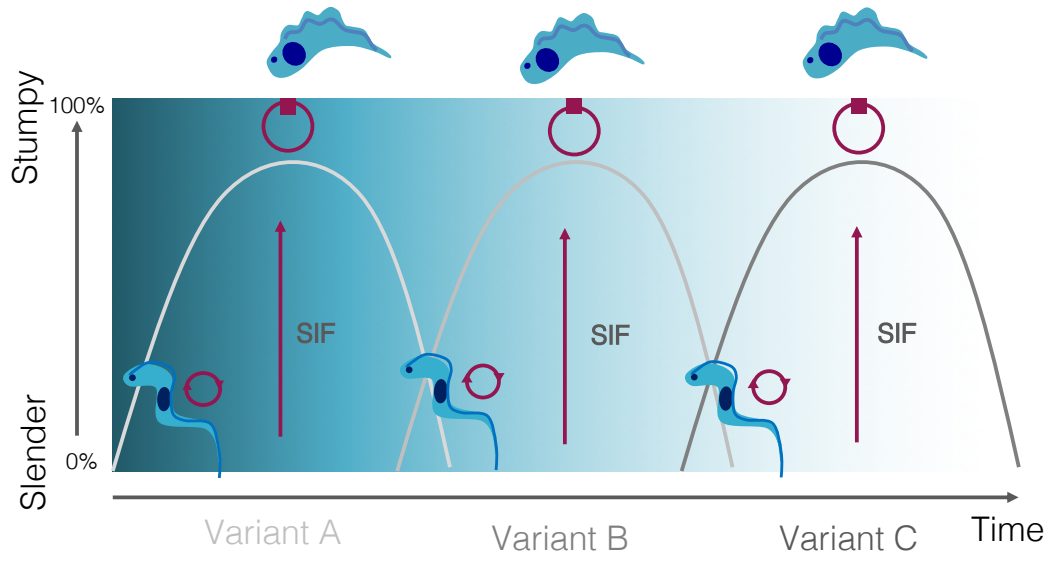


Figure 1.10 Slender to stumpy differentiation influences infection dynamics alongside antigenic variation. Proliferative long slender cells establish infection. As the parasite density increases, SIF accumulates and, upon reaching a critical density, parasites differentiate to transmission-competent, quiescent short stumpy forms. Cellular quiescence and immune recognition contribute to the fall in parasitaemia. Parasitaemia rises again as slender cells, which express an antigenically distinct VSG, begin to proliferate. Figure adapted from (MacGregor et al., 2012).

1.3.2 Morphological, physiological and gene expression changes during stumpy formation

Stumpy cells are the only bloodstream form capable of surviving within the harsh environment that the tsetse midgut presents. Accordingly, differentiation to stumpy forms is accompanied by a number of changes that preadapt the parasite for survival upon uptake by the tsetse fly (Rico et al., 2013). Following reception of SIF, and commitment to life cycle progression, stumpy cells assume their eponymous cell morphology. The cell shortens and broadens, the flagellum decreases in length, now no longer free, the flagellar pocket expands and a pronounced undulating membrane develops (Bruce et al., 1912). Intracellular remodelling of the receptor-mediated endocytic pathway furthermore translocates the single expanded lysosome to the posterior of the cell, with the lysosome sitting anterior to the nucleus (Brown et al., 1972; Vanhollebeke et al., 2010). The rate of endocytosis itself is increased in stumpy cells, allowing the cells to more rapidly clear surface bound IgG than their slender counterparts, and therefore resist complement mediated lysis until transmission occurs (Engstler et al., 2007; McLintock et al., 1993).

For the stumpy cells which have avoided antibody killing within the mammalian bloodstream for long enough to be ingested by a tsetse fly, the challenge for survival does not end here. The tsetse fly midgut is a substantially colder (~17 degrees Celsius) environment replete with its own humoral immune defences, such as antimicrobial peptides (Dyer et al., 2013). Whilst slender cells are particularly sensitive to such conditions, stumpy cells demonstrate an increased tolerance to mild pH changes and trypsin treatment (Nolan et al., 2000; Rolin et al., 1996). In fact, exposure of slender pleomorphic trypanosomes to mild acidic conditions (pH5.5) for two hours at 37 degrees Celsius has been shown to accelerate slender to stumpy differentiation through an intermediate form and furthermore permits differentiation to procyclic forms at 27 degrees Celsius in the absence of citrate/*cis*-aconitate (CCA) (Rolin et al., 1998; Szoor et al., 2013). Similarly, treatment of monomorphic *T. b. gambiense* cultures with trypsin at 27 degrees Celsius stimulated differentiation to procyclic forms (Yabu and Takayanagi, 1988) and a similar observation was made with *T. brucei* (Szoor et al., 2013).

Whilst the mammalian bloodstream is rich in glucose, the metabolite is of scarce abundance within the tsetse fly (Geigy et al., 1959). Instead the insects preferentially metabolise amino acids, in particular *L*-proline (Bursell, 1963; Mantilla et al., 2017). As such, blood parasites, which produced ATP through the glycolysis of host glucose largely within the glycosome (Opperdoes and Borst, 1977), must undergo extensive metabolic reprogramming to exploit the new nutritional environment and establish infection within the insect vector. Indeed, it has been shown that a full proline catabolic pathway is essential for trypanosomes to survive within the tsetse fly midgut (Mantilla et al., 2017). Consistent with their role in pre-adaptation to the insect life-cycle stage, stumpy cells upregulate genes involved in mitochondrial respiration (Brown et al., 1973; Capewell et al., 2013b; Michelotti and Hajduk, 1987; Tyler et al., 1997) and also elaborate the mitochondrion which can be visualised by stains such as Mitotracker® (Molecular Probes™) (Silvester et al., 2017; Vassella et al., 1997b).

Mitochondrial genes are not the only transcripts upregulated upon differentiation to the stumpy form. Stumpy cells, unlike slender forms, highly express the Proteins Associated with Differentiation (PAD) family of surface carboxylate transporters (Dean et al., 2009). Upregulated and relocated to the cell-surface from the flagellar pocket during cold-shock, PAD2 is responsible for perception of the CCA procyclic differentiation signal (Czichos et al., 1986; Dean et al., 2009; Engstler and Boshart, 2004). PAD1 is only expressed on the surface of stumpy cells and has become the gold standard of stumpy form identification. Repression of PAD1 expression in slender cells has been shown to be controlled by two regulatory sequences within the genes 3' untranslated region (UTR) (MacGregor and Matthews, 2012).

Further genes demonstrated to be consistently upregulated in stumpy forms include those that encode for: ESAG9, a rare ES component only seldom found within the VSG polycistronic transcription units (Barnwell et al., 2010; Hertz-Fowler et al., 2008); MSP-B, involved in cleavage and shedding of VSGs upon differentiation to procyclic forms (LaCount et al., 2003); TbPIP39, a glycosomal DxDxT phosphatase involved in procyclic differentiation (Szoor et al., 2010); and

NT10, a purine nucleoside transporter (Capewell et al., 2013b; Sanchez et al., 2004; Spoerri et al., 2007). Predictably, genes encoding glycolytic, cytoskeletal and cell cycle progression components were all substantially downregulated compared to slender cells (Capewell et al., 2013b; Jensen et al., 2009; Kabani et al., 2009).

1.3.3 Life cycle competency

Cells competent for the slender to stumpy transition, for example the East African Trypanosomiasis Research Organisation (EATRO) 1125 strain (http://tryps.rockefeller.edu/DocumentsGlobal/lineage_antat1.pdf), are termed 'pleomorphic'. In the absence of tsetse fly transmission, and through extensive laboratory passage at high density however, rapidly proliferating and highly virulent bloodstream from parasites that have lost the capacity to differentiate in a density-dependant manner into cell-cycle arrested stumpy forms are selected for (Ashcroft, 1960). These laboratory-adapted parasites are termed 'monomorphic'. Monomorphic trypanosomes of the Lister 427 strain (http://tryps.rockefeller.edu/DocumentsGlobal/lineage_Lister427.pdf) have historically been preferentially investigated since they generally show relative antigenic stability compared to pleomorphic trypanosomes and grow to high densities *in vitro*, thus easing their characterisation (Cross, 1996). Though monomorphic cells themselves are aberrant in their responsiveness to SIF, conditioned medium prepared from monomorphs is capable of efficiently inducing cell cycle arrest and differentiation to stumpy forms in SIF responsive pleomorphic cells (Vassella et al., 1997a). Thus, monomorphic cells must be capable of producing SIF but extensive passage has selected for either reduced reception or transduction of the differentiation signal (Silvester et al., 2017; Vassella et al., 1997a).

1.3.4 The differentiation signal

The identity of SIF has thus far remained elusive, though it has been characterised as a small (≤500Da), soluble and heat stable entity which accumulates as parasite density increases (Vassella et al., 1997a). The cyclic 3', 5'-adenosine

monophosphate (cAMP) signalling pathway attracted attention following the observations that intracellular levels of cAMP fluctuated during *in vivo* differentiation (Strickler and Patton, 1975) and increased over three-fold at peak parasitaemia (Mancini and Patton, 1981). Exposure of both pleomorphic and monomorphic cell cultures to the membrane-permeable cAMP analogue 8-(4-chlorophenylthio)-cAMP (8-pCPTcAMP) caused the induction of physiologically relevant differentiation events such as rapid G0/G1 arrest, morphological transformation and mitochondrial elaboration (Breidbach et al., 2002; Vassella et al., 1997a). Intracellular levels of cAMP furthermore increased 2-3 fold upon exposure to conditioned medium (Vassella et al., 1997a). Despite such evidence, the trypanosome genome lacked any annotated cAMP-dependant effector proteins, and finally Laxman et al. were able to demonstrate that the results from the prior experiments were due to the activity of a hydrolysis product of 8-pCPTcAMP, and not 8-pCPTcAMP itself (Laxman et al., 2006).

1.3.5 Quorum sensing signalling pathway

Whilst not the natural inducer of slender to stumpy differentiation, 8-pCPTcAMP has nevertheless proved a useful molecular tool. Combining 8-pCPTcAMP selection and RNA Interference Target Sequencing (RITseq) (Alsford et al., 2011), Mony et al. identified an extensive cohort of molecules which together formed the SIF signalling pathway (Figure 1.11a) (Mony et al., 2014). Briefly, an inducible RNAi library providing 5-fold *T. brucei* genome coverage (Morris et al., 2002) was transfected into monomorphic *T. brucei* cells. Exploiting monomorphic cells' potential to undergo cell cycle arrest and become 'stumpy-like' in response to 8-pCPTcAMP/AMP (Breidbach et al., 2002), library cell lines were grown in the presence of the SIF mimic in either induced or non-induced conditions. Uninduced populations, where the signalling pathway remained intact, rapidly underwent cell cycle arrest. However, where a gene involved in the slender to stumpy differentiation pathway was inducibly knocked down, the cells became differentiation defective and continued to proliferate in the presence of 8-pCPTcAMP/AMP. RNAi cassette insert DNA from the resistant populations was subject to Ion Torrent Deep Sequencing and

the reads mapped back to the *T. brucei* genome (Mony et al., 2014). Groups of genes identified in the genome-wide RNAi screen formed distinct positions within the predicted signalling pathway, from purine metabolism, to protein kinase and phosphatase signal transducers through to RNA-binding effector molecules (Figure 1.11b) (Mony et al., 2014). The majority of the genes were validated as having physiological significance in stumpy formation *in vivo*. The production of a chemical signal molecule upon reaching a critical cell density, and the reception of the signal through a cascade of signalling and effector molecules, is reminiscent of bacterial quorum-sensing (QS) (Bassler, 2002; Mony et al., 2014; Mony and Matthews, 2015). RITseq has furthermore been responsible for the identification of genes involved in antitrypanosomal drug resistance and action (Alsford et al., 2012), essential mitochondrial genes (Mbang-Benet et al., 2015) and genes essential for normal trypanosome growth and development (Alsford et al., 2011).

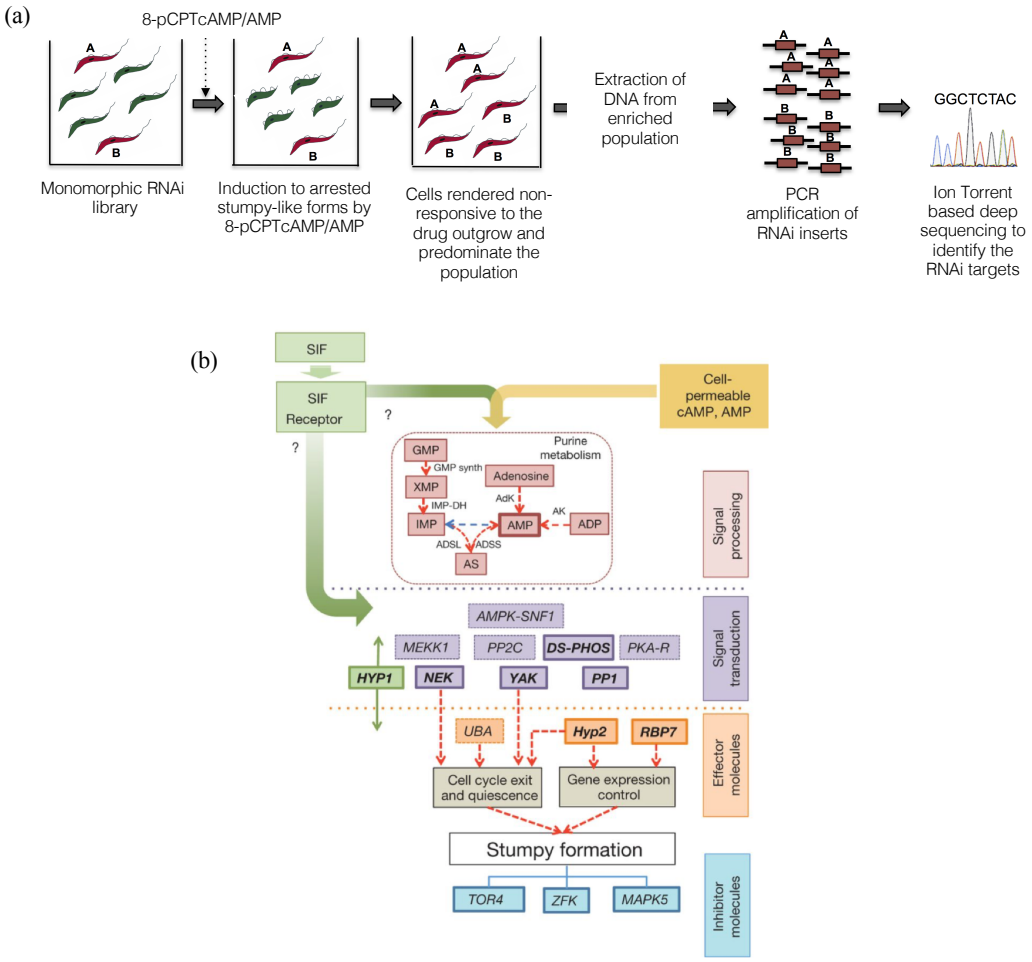


Figure 1.11 Identification of the quorum-sensing signalling pathway. (a) The genome-wide RNAi 8-pCPTcAMP/AMP screen workflow. A monomorphic RNAi library was treated with the SIF mimic, 8-pCPTcAMP. Upon exposure, uninduced cells arrested. However, cells harbouring an RNAi fragment which knocked down a component of the stumpy formation signalling pathway were rendered resistant to the signal upon induction and continued to proliferate. Amplified RNAi inserts were subjected to Ion Torrent Sequencing. **(b)** The genes identified by Ion Torrent Sequencing were ordered into a theoretical pathway. Genes highlighted in bold were experimentally validated *in vivo*. Figures from (Mony et al., 2014; Mony and Matthews, 2015).

McDonald et al. (2018) recently utilised an extragenic suppression approach to elucidate the positioning of, and dependency relationships between, a number of the stumpy formation components with respect to each other. Differentiation

incompetent RNA Binding Protein 7 (RBP7), YAK and MEKK1 null mutants were studied for their differentiation competency *in vivo* following pairwise inducible ectopic overexpression of either RBP7, NEK17 or Protein Phosphatase 1-6 (PP1-6) or the inducible gene silencing of the differentiation inhibitor TbTOR4 (McDonald et al., 2018). Cells were assessed for their PAD1 expression, morphology, cell cycle status and competency for synchronous differentiation to procyclic forms. The study revealed complex non-linear relationships between the components (Figure 1.12) and furthermore that cell cycle arrest was not always a prelude to stumpy formation, as demonstrated by low PAD1 expression in an arrested MEKK1 null mutant lacking TbTOR4 (McDonald et al., 2018).

It has been demonstrated that the density-dependent quorum-sensing signalling pathway is not the only means by which life-cycle progression may be mediated. Transcriptional attenuation of the active VSG ES in monomorphic cell populations was shown to be followed by an accumulation of (reversibly) G1 arrested cells expressing PAD1, which were somewhat more efficient at differentiating to procyclic forms when stimulated with CCA (Batram et al., 2014). Simultaneous ectopic overexpression of two VSGs in biologically relevant pleomorphic cells, however, produced two phenotypically distinct populations (Zimmermann et al., 2017): those which were irreversibly arrested, and those which remained proliferative. In the arrested populations, overexpression of the second VSG had caused transcriptional attenuation of the active VSG ES. These cells had additionally accumulated in G1, upregulated PAD1 expression and were capable of CCA-induced synchronous differentiation to procyclic forms (which themselves were capable of establishing a tsetse infection). Thus, the VSG ES transcriptional attenuation had elicited life-cycle progression in a density independent manner. Contrastingly, in the proliferative population, though the endogenous active VSG had been silenced following ectopic overexpression of the second VSG the ES remained to some extent transcriptionally active (Zimmermann et al., 2017). The authors postulated that such an outcome was evidence of an ‘escape mechanism’ utilised by the parasites whereby, in the instance of a failed VSG switch and complete transcriptional shut-down of the VSG ES, the cells could differentiate to transmission competent stumpy forms. If this failed switch did not cause complete

transcriptional shut down of the VSG ES, as in the second population of proliferative VSG overexpressors, the cells could continue to replicate as slender forms (Silvester et al., 2017; Zimmermann et al., 2017).

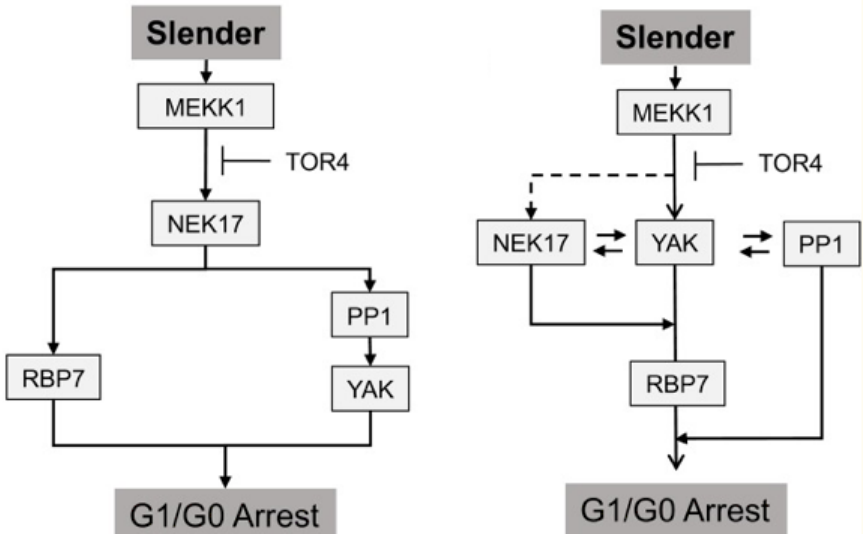


Figure 1.12 Two potential quorum-sensing hierarchy models. Two interpretations of the knockdown and overexpression data generated by (McDonald et al., 2018). Alternative models to explain the data may exist. The model on the left consists of two non-redundant branches, one YAK- and one RBP7-dependant, leading to cell-cycle arrest and stumpy formation. The model on the right is linear, and the activities of NEK17 and PP1-6 are dependent on YAK. The dotted line indicated the potential for NEK17 phosphorylation, either directly or indirectly, by MEKK1. Figure from (McDonald et al., 2018).

1.4 Antigen switch frequency and stumpy differentiation capacity: a link?

Historically, VSG switch studies have estimated VSG switch frequencies of 10^{-7} to 10^{-5} switches/cell/generation, a rate that reflects the background level of mutations in other eukaryotes (Doyle et al., 1980; Lamont et al., 1986; Myler et al., 1985; Turner, 1997; Watkins, 1964). Such studies, crucially, made use of laboratory-adapted monomorphic trypanosomes that had been subjected to continued syringe passage through rodent hosts, thus removing the parasite several thousand

generations from transmission. To date, as discussed in Section 1.3.3, monomorphic trypanosomes are still preferentially investigated in VSG studies owing to their relative antigenic stability, ease of characterisation and annotated VSGnome (Cross et al., 2014).

Populations of syringe-passaged monomorphic clones are typically antigenically homogenous, with over 99% of the population expressing the same dominant VSG. However, moderate to high levels of antigenic heterogeneity are observed in populations of life-cycle competent cells that have not yet undergone extensive laboratory passage, suggesting that antigenic stability may differ between pleomorphic and monomorphic cells (Barry and Emery, 1984; Doyle et al., 1980; Turner and Barry, 1989). Turner and Barry (1989) estimated the VSG switch rate of *T. b. rhodesiense* cells less than 100 generations removed from a tsetse fly to be 2-9.3x10^{-3} switches/cell/generation, a phenomenon that was shown to be general to trypanosomes and not a specific line. This estimated switch rate, however, was a summation of a series of immunofluorescence experiments designed to test the switch rate between specific combinations of VSGs, rather than the populations' overall switch frequency. It was proposed that, had analysis not been restricted to a set of six VSGs, the estimated switch rate would have in fact been higher (Turner and Barry, 1989). Switch rates were highly variable (from 1.9x10^{-6} switches/cell/generation to 6.9x10^{-3} switches/cell/generation) and dependent on the VSG that was being switched to (Turner and Barry, 1989). The observation that recently isolated trypanosomes could switch their expressed VSG at a rate 66,000-fold times greater than cells subjected to laboratory passage was corroborated by (Turner, 1997). Excluding trypanosome genotype and methodological differences between studies as explanations for the discrepancies in published VSG switch rates, Turner used an *in vivo* method conceptually based on Luria and Delbrück's fluctuation analysis to show that rates of antigenic variation were significantly lower in laboratory-adapted lines compared to those recently transmitted by tsetse flies (Turner, 1997). Briefly, two cloned *T. brucei* lines were laboratory-adapted by extensive syringe passage through mice every 2-3 days for a period of three months. Mice were injected with no more than three of the syringe passaged clones to allow for inoculation of a homogenous population with regard to VSG expression.

Infections were allowed to develop for 6 days before the mice were exsanguinated and non-VSG-switched cells killed by complement-mediated lysis. Groups of mice were then inoculated with a known number of the live and/or killed cells. Mice were observed for their developing parasitaemia over 14 days, with the VSG switch rate finally being estimated by the following

equation:

$$Switch\ Rate = \frac{0.693(-logePo)}{N}$$

where P_0 is the proportion of mice that did not develop infection and N is the number of live/dead cells injected into each mouse (Luria and Delbruck, 1943; Turner, 1997). The switch rates of the two passaged clones were calculated by this method, and estimated at $3x10^{-6}$ switches/cell/generation and $1x10^{-4}$ switches/cell/generation, respectively. The VSG switch rates of 4 subpopulations were determined after the former clone was subject to cyclical transmission through a tsetse fly. VSG switch rates post cyclical transmission were variable, but consistently higher than those measured before tsetse fly transmission. Three of the four populations switched their expressed VSG at the rate of $2x10^{-3}$ switches/cell/generation, the fourth considerably less so at $7x10^{-5}$ switches/cell/generation. Since the reduction in VSG switch rate could be readily reversed upon cyclical transmission, it was postulated that the change was unlikely to be due to mutation (Turner, 1997).

Though with considerable variability, both the above studies have demonstrated that recently-isolated, life-cycle competent (pleomorphic) trypanosomes switch their expressed VSG at a significantly higher rate than populations of cells which have been subject to extensive laboratory passage. It therefore seems that stumpy differentiation capacity and antigen switch frequency could be linked.

1.4.1 The evidence for a link

Few studies have formally addressed the evidence for, nor hypothesised the reasons for, a direct link between antigen switch frequency and stumpy

differentiation capacity. Using the same low switching trypanosome lines that a) have now been passaged as monomorphic cells for decades and b) have likely been shared between many laboratories for most published VSG switching studies, "monomorphic trypanosomes switch their expressed VSG at a significantly lower frequency than pleomorphic cells" has become a dogma of trypanosome antigenic variation. Do all monomorphic strains behave this way? What about cells that have only recently lost their pleomorphism? Can higher VSG switch rates be selected for?

Though in the context of a link between switch *frequency* and stumpy differentiation *capacity* evidence may be lacking, there is abundant evidence to suggest slender to stumpy differentiation and antigenic variation themselves are coordinated. Development to stumpy forms is accompanied by silencing of the active VSG ES (Amiguet-Vercher et al., 2004) and, as explained in detail in Section 1.3.5, full transcriptional attenuation of the active VSG ES has furthermore been shown to elicit life cycle progression in pleomorphic cells (Batram et al., 2014; Zimmermann et al., 2017). RNAi mediated knock down of either VSG 221 or VSG 117 (Lister 427 variants) has additionally been shown to induce cell cycle arrest, however this is both reversible and at precytokenesis, rather than in G1 (Sheader et al., 2005; Smith et al., 2009).

Dreesen et al. (2007) have hypothesised that telomere length and antigen switch frequency may be inversely related. When the telomeres are short, large stochastic terminal deletions of the active BES fall within the subtelomeric 70 bp repeat, and thus cause a break-induced replication (BIR) mediated VSG switch. Conversely, when the telomeres are long, the deletions fall within the telomere region itself and instead cause a further lengthening of the telomere and no switch (Dreesen and Cross, 2008; Dreesen et al., 2007). Slender to stumpy differentiation became of relevance when it was observed that recently isolated pleomorphic cells, which switched their expressed VSG at a high rate, had significantly shorter telomeres that those of extensively propagated monomorphic cells which switched their expressed VSG at a low frequency (Dreesen and Cross, 2008). Monomorphic telomerase reverse transcriptase null mutants ($TERT^{-/-}$) incapable of lengthening their telomeres were shown to have a significantly increased switch frequency

compared to wild-type cells, in particular using GC mechanisms (Hovel-Miner et al., 2012). However, it was not proven whether this increase in switch frequency was directly a result of shortened telomeres, or whether it was the result of a telomere capping defect caused by the deletion of *TERT*, as has previously been documented in the budding yeast *Kluyveromyces lactis* (Hovel-Miner et al., 2012).

A change in antigenic variation upon culture adaptation has not only been observed in *T. brucei*. Serial blood passage (SBP) of *Plasmodium chabaudi chabaudi* enhances parasite virulence and furthermore selects for expression of a dominant *cir* gene (the *P. c. chabaudi*-specific *pir* multigene family which encodes for variant surface antigens, VSA) (Spence et al., 2013). RNAseq analysis of SBP parasites isolated following mosquito transmission found that vector transmission was associated with modified expression of the *cir* multigene family, most likely by epigenetic reprogramming, with 57% of the *cir* genes upregulated across the population (Spence et al., 2015; Spence et al., 2013). A decrease in VSA expression diversity upon SBP, and the subsequent upregulation of the composition and frequency of transcripts following vector transmission, was also observed in human infections with cultured *P. falciparum* (Peters et al., 2002). Parasites which had recently gone through a sexual cycle in the mosquito vector switched their *var* genes at an average of 16% per generation, considerably greater than an *in vitro* grown line of *P. falciparum*, which was shown to switch at a rate of 2% per generation (Peters et al., 2002; Roberts et al., 1992). The pili of the bacterial pathogen *Neisseria gonorrhoeae* are important virulence factors and can be switched to antigenically distinct alternatives via a DNA recombination pathway similar to that of gene conversion in trypanosomes (Serkin and Seifert, 2000). Whilst pili expression is relatively stable *in vitro,* multiple pili variants are recovered following human infection (Koomey, 1997; Seifert et al., 1994). The frequency of pilin antigenic variation was shown to be influenced *in vitro* by the availability of iron, whereby cells cultured in iron-limiting conditions exhibited switch rates five-fold greater than cultures grown in iron-replete conditions (Serkin and Seifert, 2000). Finally, as few as 10 *in vitro* passages of the spirochete *Borrelia burgdorferi*, the causative agent of Lyme disease, selects for low-infectivity parasites missing the 28kb linear plasmid which stores their antigenic variation machinery (Koomey, 1997).

1.5 Aims

The objective of this study was to ask if there is a direct relationship between *T. brucei* stumpy formation capacity and VSG antigen switch frequency. To address if loss of stumpy formation capacity was associated with a concomitant reduction in VSG switch frequency, the following approaches were applied:

i) Generation of a sensitive *in vitro* flow cytometry-based assay capable of detecting VSG switch events in populations of pleomorphic and inducibly monomorphic cells.

ii) Utilisation of a targeted deep sequencing approach, VSGseq, to determine expressed VSG diversity in populations of pleomorphic and inducibly monomorphic cells.

iii) Analysis of selected monomorphs in the flow-cytometry based VSG switch assay and by whole genome and transcriptome sequencing.

The findings of this study will contribute towards our understanding of the complex interplay between antigenic variation and differentiation, and help us to understand how, together, they influence the dynamics of chronic trypanosome infection.

2 : MATERIALS AND METHODS

2.1 Trypanosome strains

The pleomorphic *T. brucei* AnTat1.1 (EATRO1125) 90:13 trypanosome strain (Engstler and Boshart, 2004) was used for generating the inducibly monomorphic and selected monomorphic cell lines in this study. The *T. brucei* AnTat1.1 90:13 strain contains an integrated T7 RNA polymerase and Tet repressor to facilitate inducible genetic manipulations.

The wild-type monomorphic trypanosomes used in Chapters 3 and 5 of this study were of the *T. brucei* Lister 427 strain and the stabilates were acquired from the Schnaufer Lab (University of Edinburgh, UK). The VSG^{up} cell line used in Chapter 3 was also based on the *T. brucei* Lister 427 strain, however the stabilates were acquired from the Horn Lab (University of Dundee, UK).

2.2 Trypanosome culture

Bloodstream form trypanosome cultures were maintained in HMI-9 medium supplemented with 20% foetal calf serum (Gibco), 100U/ml penicillin and streptomycin and the appropriate selective drug(s) (Table 2.1) at 37°C and 5% CO_2 (Hirumi and Hirumi, 1989). Pleomorphic strains were maintained below densities of $1x10^6$ cells/ml and monomorphic strains below densities of $2x10^6$ cells/ml. To generate parasite stabilates, $2x10^6$ cells were centrifuged at 1500g and resuspended in 500µl of HMI-9/14% glycerol. Cryovials were initially contained within a polystyrene box when moved to -80°C to prevent ice crystal formation. Cryovials were removed from polystyrene boxes after 24 hours and stored long term at -80°C and in liquid nitrogen.

To thaw a stabilate, the cryovial was incubated for 1 minute at 37°C and the contents added to 2.5ml fresh HMI-9. This was centrifuged at 1500g for 5 minutes to remove the glycerol and the pellet resuspended in 10ml fresh HMI-9. Selective drug was added to the culture once it was clear that the parasites had survived the freeze/thaw cycle. If a bloodstock was thawed the same process was followed, except that the 10ml flask was stored in an upright position within the incubator. This

caused the red blood cells to settle to the bottom of the flask and the next day the top layer of HMI-9 could be carefully decanted and transferred to a new flask. This process was continued until red blood cells were no longer visible in the culture.

Cells from culture volumes of more than 2ml were counted using a Beckman Z2 Coulter particle counter and size analyser (250µl of cells were required/count). For cultures of ≤ 2ml, 10µl of cells were removed and counted on a haemocytometer.

Table 2.1: Selective drug concentrations for the *T. brucei* cell lines used in this study. All drugs were acquired from InvivoGen.

CELL LINE	SELECTIVE DRUG(S)	CONCENTRATION (µg/ml)
AnTat1.1 90:13	Hygromycin	0.5
	G418	2.5
AnTat1.1 90:13 pALC14	Hygromycin	0.5
	G418	2.5
	Puromycin	0.5
AnTat1.1 90:13 pALC14 pLF12 eGFP_BLA	Hygromycin	0.5
	G418	2.5
	Puromycin	0.5
	Blasticidin	10
Lister 427 VSG^{UP}	Phleomycin	20
	Hygromycin	2.5
	Puromycin	2

2.3 cAMP growth assays

To test differentiation capacity *in vitro,* cells were exposed to the membrane-permeable cAMP analogue, 8-(4-chlorophenylthio)-cAMP (8-pCPTcAMP). Incubation with 8-pCPTcAMP causes rapid G1 arrest in pleomorphic cells (Laxman et al., 2006), but knockdown of a gene involved in the stumpy formation pathway would be expected to prevent the cells from G1 arrest as cells remain slender and proliferative (Mony et al., 2014).

Four conditions were tested in triplicate for each cell line: -DOX/-8-pCPTcAMP; -DOX/+8-pCPTcAMP; +DOX/-8-pCPTcAMP and +DOX/+8-

pCPTcAMP. On day 0, the cells were seeded at $1x10^5$ cells/ml in a total volume of 2ml in a 24 well plate. RNAi was induced in the +DOX replicates on day 0 by the addition of 1μg/ml doxycycline dissolved in DMSO. The 8-pCPTcAMP was not added to the +8-pCPTcAMP samples until 24 hours later to allow sufficient time for transcript knockdown. Cell were counted at 24 hour intervals as described in Section 2.2. Cells were diluted back to ~$1x10^5$ cells/ml, when required, to prevent growth above $1x10^6$ cells/ml and thus, SIF-mediated arrest. If a population was diluted, doxycycline or 8-pCPTcAMP was added to the fresh HMI-9, at the appropriate concentration. Cumulative parasite growth was plotted using GraphPad Prism 7.0d software.

2.4 Trypanosome infections in mice

All mice used in this study were MF1 strain females and housed in the MARCH Building, University of Edinburgh. MF1 mice are used due to their relatively placid nature and large size, thus allowing more material to be collected on the final day of infection. MF1 mice are, however, outbred and this can lead to experimental variation between infections. All the infections were performed by myself under the PIL number I42D10074, except for those presented in Figure 3.2 which were kindly performed by Julie Young. All infections were performed according to the UK Home Office and Project Licence (PPL number 604373) regulations.

For the generation of slender form bloodstocks, 200μl of HMI-9 containing 10,000 parasites were injected intraperitoneally into mice. The parasitaemia was monitored daily from day 3 onwards by observing for parasites in blood collected on a slide from a tail snip. The parasitaemia was estimated using the rapid "matching" method described by Herbert and Lumsden (1976). Bloodstock infections were terminated when the parasitaemia reached 3.1-$6.3x10^7$ slender cells/ml. Terminal anaesthesia was induced by placing the mouse in a holding chamber that was then flooded with isoflurane. Once the mouse no longer showed a reflex response (as judged by squeezing a paw firmly with tweezers) it was removed from the holding

chamber and placed on its back with its head in a small cone delivering isoflurane. Whole blood was extracted by cardiac puncture into a 2ml syringe containing 200µl of 2% citrate to prevent the blood from clotting. After cardiac puncture, experimental mice were euthanised by cervical dislocation.

For analysing trypanosome differentiation competence over the course of an infection, infections were established from blood stocks and not *in vitro* grown parasites. The number of cells injected varied depending of the virulence of the strain, but ranged from 12,500-750,000 cells diluted in HMI-9 to a total volume of 200µl. The parasitaemia was monitored daily as previously described. For tracking cell cycle status and morphology across the infection time-course, a large drop of tail blood was smeared across a microscopy slide that was then air-dried and fixed in ice-cold methanol. The slides were stored in methanol at -20^{o}C until further use. Infections were terminated either when the population had differentiated to stumpy forms or, in the case of parasites that had lost their differentiation capacity and remained slender, before the parasitaemia became lethal. Blood was extracted by cardiac puncture as before.

Parasites were purified from whole blood by passing the blood through a DEAE-C anion exchange column (Whatman) at pH7.8 with PSG buffer (see below). At pH7.8, blood is retained on the column whilst the parasites are collected in the flow-through. Blood can be washed from the column by flushing the resin with dH_2O.

<u>PSG Buffer</u>

44mM NaCl

83mM *D*-glucose

3mM $NaH_2PO_4.2H_2O$

67mM Na_2HPO_4

pH 7.8

2.5 DNA cloning

2.5.1 Genomic DNA extraction

Genomic DNA (gDNA) was extracted from $5x10^7$-$1.5x10^8$ parasites. Trypanosomes were pelleted from *in vitro* culture or DEAE-C column flow-through by centrifugation at 1500g for 5 minutes. The pellet was washed three times in sterile 1X PBS and, after the last wash, the PBS was removed and the pellet was resuspended in 150µl TELT lysis buffer (see below) using a wide bore tip. After 5 minutes incubation at room temperature (great care was taken not to incubate for >5 minutes), 150µl (1:1 v/v) of water equilibrated phenol/chloroform (Sigma-Aldrich) was added. The tube was slowly shaken by hand for 5 minutes before being centrifuged at 15000g for 10 minutes. Next, the upper phase was removed and transferred to a new tube without disturbing the proteinaceous interface. The phenol/chloroform extraction was then repeated as before, isolating the upper phase. After the second extraction, 300µl of 100% ethanol was added to the upper phase to remove traces of phenol. The tube was mixed by hand for 15 seconds then incubated for 5 minutes at room temperature. The tube was then centrifuged at 15000g for 10 minutes. The supernatant was removed and the pellet washed once more with 100% ethanol before being centrifuged at 15000g for 5 minutes. The pellet was dried on a bench at room temperature and then resuspended in 20-50µl dH_2O containing 20µg/ml RNase. The DNA content was quantified using a Nanodrop™ Spectrophotometer (Thermo Fisher Scientific). All DNA samples (except for those used for whole genome sequencing, see Section 2.18) were stored at -20°C.

TELT Buffer

50mM Tris-HCl, pH8.0

62.5mM EDTA, pH9.0

2.5M LiCl

4% v/v Triton X100

If further DNA purification was required, the DNA was precipitated overnight at -80ºC in 0.1 volumes 3M sodium acetate and 2.2 volumes ice-cold 100% ethanol. Samples were centrifuged at 15000g for 30 minutes, then washed in ice-cold 70% ethanol and centrifuged at 15000g for 20 minutes. The pellet was dried and resuspended as before.

2.5.2 PCR

PCR reactions were performed in 0.2ml PCR tubes (Corning) in a Biometra thermocycler. If the purpose of the PCR reaction was to assess the presence of a gene or DNA fragment, the GoTaq® G2 DNA Polymerase (Promega) was used according to the manufacturer's instructions. If the PCR product was to be used further for cloning or for making a DNA probe, the proof-reading Phusion® High-Fidelity DNA polymerase was used according to manufacturer's instructions. All primers used in this study were produced by Sigma-Aldrich and designed using Primer3 software (with the exception of the PAC Probe_FWD and PAC Probe_RVS primers, which were designed by Dr Federico Rojas) (Koressaar and Remm, 2007; Untergasser et al., 2012). Primer T_m was calculated using the following equation, and the PCR annealing step run at 2-4ºC below the lowest T_m of the primer pair:

$$T_m = 4℃\, x\, (number\ of\ G\&C\ bases\ in\ primer) + 2℃\, x\, (number\ of\ A\&T\ bases\ in\ primer)$$

Table 2.2: Primer sequences used in this study. All sequences are written in the 5' to 3' direction. Restriction enzyme sites are underlined.

Primer Name	Notes	Sequence
SPLICED LEADER_FWD	Used in VSG-specific RT-PCR to amplify conserved 5' and 3' ends of all VSG mRNAs	ACAGTTTCTGTACTATATTG
14MER_RVS		TAAGCAGTGTTAAAATATATC
PAC Probe_FWD	Amplifies puromycin N-acetyltransferase gene	TTAGCGCGCTAAACAGCAAGGTCTTCTGAAATT CATG
PAC Probe_RVS		TTAAGGCCTTCAGGCACCGGGCTTGC
5' OUT_FWD	Used in conjunction with PAC Probe_RVS AND 3' OUT_RVS primers to verify corect integration of the *PAC* gene in the VSG 221 ES	TAAGCACTAAGAGCAGTAATAAGTAT
3' OUT_RVS	Used in conjunction with PAC Probe_FWD AND 5 'OUT_FWD primers to verify corect integration of the *PAC* gene in the VSG 221 ES	TAAGCAGCAATATTCCACATCATCA
5' VSG 221_RVS	Used in conjunction with PAC Probe_FWD primer to verify corect integration of the *PAC* gene in the VSG 221 ES	TAAGCAAGAACTTGGGCTAGGACCAAGA
HYP2_FWD	qRT-PCR primer	TAAGCATTGGCCAAACTGTGATGTCG
HYP2_RVS		TAAGCACATGCCTGAGATGAACGGTG
NEK_FWD	qRT-PCR primer	TAAGCAGATGGGGAGAGTATGCGTGA
NEK_RVS		TAAGCAAAGGTCGCTGCTTAGGATCA
DYRK_FWD	qRT-PCR primer	TAAGCACGCGCGTACAGAATTCTCAA
DYRK_RVS		TAAGCAGGTGCCCTATGTAGCGTTTG
ZFP3_FWD	qRT-PCR primer	CAGGGGAAACGCAAAACTAA
ZFP3_RVS		TGTCACCCCAACTGCATTCT
NEK21_FWD	qRT-PCR primer	GCGGTATCTCACTTTGGCTG
NEK21_RVS		TAAGCAGATTGTGTCGAGTGGCATCC
TB927.5.4020_FWD	qRT-PCR primer	TAAGCAGAGCACGGAGGGAATGAAAC
TB927.5.4020_RVS		TAAGCATTGCATGTACAGCCCTCTCA
ZC3H20_FWD	qRT-PCR primer	TAAGCACTGCCTCCGGTACCTCTATG
ZC3H20_RVS		TAAGCACATCTCCAAGTGTCACGCTG
a1800gc1801a_FWD	Mutates two bases within the I*Sce*-I restriction site in the pLF12 eGFP_BLA plasmid	CTCACCATATTACCCTTCTATCCCTAAAGCTTC GTTGCAGTTG
a1800gc1801a_RVS		CAACTGCAACGAAGCTTTAGGGATAGAAGGGTA ATATGGTGAG
Lister_70.11_FWD	Amplify puromycin resistance gene and UTRs from pLF12 eGFP_Puro plasmid. FWD and RVS primers add VSG 221 ES integration sites. FWD primer also adds I-*Sce*I restriction site.	AATAGGAGAGTGTTGTGAGTGTGTGCTTACCAA TATTATAATAATGATAGTAACGACCAATAGGGAT AACAGGGTAATGGCACAGCAAGGTCTTCTG
Lister_70.11_RVS		CAAATCCATTATACTCACGATTACTCTCATTGCA CACATACCATTGTCTTAACTGCATTTATTTATGG TTTGTATTCGTCGGGCTCGAATCCCCCCATTT
BLA_FWD	Amplify blasticidin resistance. HindIII restirction site	TAAGCAAAGCTTATGCCTTTGTCTCAAGAAG
BLA_RVS	BamHI restriction site	TAAGCAGGATCCTTCGAATTAGCCCTCCCAC

2.5.3 DNA agarose gel electrophoresis

DNA agarose gel electrophoresis was performed to separate and visualise DNA fragments. Except for Northern and Southern blot experiments (see Sections 2.13 and 2.15, respectively), 1% agarose gels were prepared. To prepare a small 1%

agarose gel, 1g of agarose powder (Invitrogen) was dissolved in 100ml of 1X TAE buffer (see below) by heating, or for larger 1% agarose gels, 1.5g of agarose was dissolved in 150ml of 1X TAE buffer. Once the powder had dissolved, 0.5µg/ml ethidium bromide was added and the gel poured into a cast within a gel electrophoresis tank. Once the gel had set, the tank was filled with 1X TAE buffer and the DNA samples loaded into the wells alongside a 1kB DNA ladder to allow for size estimation. The samples were electrophoresed at 120V until the DNA fragments were resolved. The DNA was visualised and photographed on a Syngene G:box UV transilluminator.

1X TAE Buffer

40mM Tris-Acetate

1mM EDTA

pH 8.0

2.5.4 Column DNA purification

When PCR amplification produced a single band of the correct size, the PCR reaction product was purified over a column using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs) according to the manufacturer's instructions. The typical final elution volume was 6µl. Products were stored at -20ºC for further use.

2.5.5 DNA purification from agarose gel

If a PCR reaction resulted in the amplification of an unspecific product in addition to the desired target, the correct DNA amplicon was purified from a DNA agarose gel using the Monarch® DNA Gel Extraction Kit (New England Biolabs) according to the manufacturer's instructions. The typical final elution volume was 6µl. Since a large proportion of DNA is lost during the gel extraction procedure, at least 3 PCR reactions were purified together and the products loaded into as few

wells as possible in the agarose gel, to ensure that enough DNA was isolated. Prior to cutting, the DNA agarose gel was exposed to long-wavelength UV light for as little time as possible to prevent DNA damage. Products were stored at -20°C until use.

2.5.6 Restriction digests

Genomic DNA or prepared plasmids were digested with the relevant restriction enzymes and optimal reaction buffer at 37°C for either 1-3 hours (NEB High-Fidelity Restriction Endonucleases only) or overnight for all other restriction enzymes, according to the manufacturer's instructions. Where a digest with 2 different restriction enzymes was required, and if the reaction buffer was the same for both enzymes, the digests were performed simultaneously. If, however, the restriction endonucleases required different reaction conditions, sequential restriction digests were performed with the DNA being purified over a column (as Section 2.5.4) between digests.

2.5.7 Ligation

Ligation reactions were performed using the Roche Rapid DNA Ligation Kit according to the manufacturer's instructions. All ligation reactions were performed in triplicate using 1:1, 3:1 and 7:1 insert DNA to vector DNA molar ratios. To calculate the nanograms of insert DNA required, the Promega Ligations Calculator was used (https://www.promega.com/a/apps/biomath/?calc=ratio). Ligation reactions were incubated at room temperature for 1-2 hours before transformation into XL-1 *Escherichia coli* competent cells (section 2.5.9).

2.5.8 Preparation of XL-1 *Escherichia coli* competent cells

To prepare stocks of XL-1 *E. coli* competent cells, 5ml of autoclaved LB broth was inoculated with 5µl of XL-1 *E. coli* cells and incubated overnight at 37°C in a shaking incubator. The next morning, the 5ml culture was divided into two and used to inoculate 2x200ml of autoclaved LB broth in 1 litre flasks. The cultures were

incubated at 37°C in a shaking incubator until the OD_{600} was between 0.4-0.5, as measured on an Eppendorf Biophotometer. The culture was then divided into 8x50ml tubes and the cells harvested by centrifugation at 1000g for 10 minutes at 15°C. Each of the pellets was carefully resuspended in 25ml of ice-cold buffer RF1 (see below) and then incubated on ice for 1 hour. The cells were again pelleted by centrifugation at 1000g for 5 minutes at 4°C and each pellet resuspended in 8ml of ice-cold buffer RF2 (see below). The competent cells were then aliquoted in 200-400µl volumes and snap frozen in liquid nitrogen before storage at -80°C. Once an aliquot of competent cells had been thawed, it was not refrozen for further use.

Buffer RF1	Buffer RF2
10mM $RbCl_2$	10mM $RbCl_2$
50mM $MnCl_2.4H_2O$	10mM 3-(*N*-morpholino)propanesulfonic acid (MOPS)
30mM CH_3CO_2K	75mM $CaCl_2$
10mM $CaCl_2$	15% (v/v glycerol)
10% (v/v) glycerol	pH6.8, autoclave
pH5.0, autoclave	

2.5.9 Transformation of XL-1 *Escherichia coli* competent cells

Competent cells were removed from storage at -80°C and thawed on ice (100µl of XL-1 competent cells were required per transfection). Approximately 2-4µl of a ligation reaction or 1µl of a plasmid preparation was added to 100µl of thawed cells in a 1.5ml Eppendorf tube and the tube gently mixed. The cells were then incubated for a further 20-30 minutes on ice. Cells were heat shocked in a 42°C water bath for 40 seconds before immediately being placed on ice for at least 2 minutes. To allow the cells to recover before selection, 800µl of autoclaved LB broth was added and the tube then incubated at 37°C for 45 minutes with occasional shaking.

The pGEM®-T Easy vector (Promega) used in this study allowed for blue/white screening for recombinant colonies. Whilst bacteria transformed with the

pGEM®-T Easy vector were recovered at 37ºC, 100µl of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 2µl of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) were spread onto a selective agar plate. The pGEM®-T Easy vector contains a sequence encoding the first 146 amino acids of β-galactosidase, within which is the multiple cloning site (MCS). The XL-1 *E. coli* cells are *lacZ*ΔM15 mutants and therefore cannot produce functional β-galactosidase. However, when transformed with an 'empty' pGEM®-T Easy vector, a functional β-galactosidase enzyme is produced by α-complementation. β-galactosidase hydrolyses X-gal and the product of this hydrolysis reaction spontaneously dimerises to produce an insoluble blue pigment (Ullmann et al., 1967; Welply et al., 1981). When a cloning fragment has been successfully ligated into the MCS of the vector the *lacZ* gene is disrupted and thus no functional β-galactosidase is produced. Therefore, bacteria that have been transformed with an empty vector are visualised as blue colonies, whilst those that have been transformed with a ligated vector form white colonies.

When a ligation reaction had been used for the transformation, the cells were centrifuged at 1000g for 5 minutes and the pellet then resuspended in 100µl of autoclaved LB broth. The cells were then spread onto a pre-warmed selective agar plate containing 1mg/ml ampicillin and incubated at 37ºC overnight. When a prepared plasmid had been used, the cells were not centrifuged and 100µl was taken directly from the tube and spread onto the plate instead. The next morning the plates were removed from 37ºC and singular colonies (specifically, white singular colonies when blue/white screening was performed) were selected for plasmid preparation (see Section 2.5.10).

2.5.10 Plasmid preparation

For preparation of plasmid DNA, a single bacterial colony was picked from selective agar and used to inoculate 10ml of LB broth containing 1mg/ml ampicillin. Cultures were grown overnight in a shaking incubator at 37ºC. The next morning, cultures were centrifuged at 1000g for 10 minutes and the supernatant discarded. The

pellets were resuspended in 300µl of solution I (see below) containing 100µg/ml RNase and transferred to a 1.5ml tube. To this, 300µl of solution II (see below) was added and the tube mixed several times by hand. After 5 minutes incubation, 350µl of ice-cold solution III (see below) was added to the sample. The tube was shaken by hand and incubated for a further 5 minutes on ice. The sample was then centrifuged at maximum speed for 15 minutes at 4°C, the supernatant removed to a fresh 1.5ml tube, and this centrifuged at 15000g for a further 10 minutes at 4°C. The supernatant was again transferred to a new 1.5ml tube and 650µl of ice-cold isopropanol added. The tube was mixed by inverting several times and then incubated at -20°C for a minimum of 15 minutes. The sample was then centrifuged at 15000g and 4°C for 25 minutes, before the supernatant was removed and the pellet washed in 1ml of ice-cold 70% ethanol. After a final 5 minute spin at 15000g and 4°C, the pellet was air-dried and resuspended in 50µl dH_2O.

Solution I	Solution II	Solution III
50mM Glucose	0.2M NaOH	3M Potassium acetate
25mM Tris, pH8.0	1% SDS	5M Glacial acetic acid
10mM EDTA, pH8.0		
Autoclave		

2.5.11 DNA sequencing

To confirm ligation of the correct DNA product, prepared plasmids were sent for 'in-house' sequencing at Edinburgh Genomics, The University of Edinburgh. The 'Big Dye Reaction Required' service was used, and reactions set up as follows:

1µl plasmid DNA

1µl 6mM primer

4µl dH_2O

2.5.12 DNA mutagenesis

DNA site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's protocol, with the exception of the final transformation step. Mutated plasmids were transformed into XL-1 competent cells (as in Section 2.5.9) and not XL10-Gold ultracompetent cells.

2.6 Trypanosome transfections

The same experimental procedure was followed for transfecting pleomorphic and monomorphic cell lines.

Prior to transfection, 15µg of plasmid DNA (Section 2.5.10) was linearized overnight by the relevant restriction enzyme(s) (Section 2.5.6). The restriction digest reaction was purified over a DNA purification column (Section 2.5.4) and then ethanol precipitated at -80^{o}C overnight as previously described (although two bands were produced in the digest reaction the plasmid backbone was not removed since it did not contain homologous regions to the trypanosome genome and therefore would not integrate instead of the intended DNA fragment). After centrifugation of the DNA (Section 2.5.1) the pellet was air dried and resuspended in 10µl of dH_2O.

Overall, $3x10^7$ cells were used for each transfection, including for a 'no DNA' control transfection that was performed simultaneously. The cells were centrifuged at 1500g for 5 mins and the pellet washed twice in sterile 1X PBS pre-warmed to 37^{o}C. The pellet was resuspended in 50µl of Amaxa Transfection Buffer (Amaxa Basic Parasite Nucleofector Kit 2) and 10µl of the linearized DNA. To the 'no DNA' control, 10µl of sterile dH_2O was added instead. The cells and DNA/ dH_2O were mixed well, transferred to an Amaxa transfection cuvette and electroporated in an Amaxa Nucleofector II using the Z-001 programme. The electroporated cells were immediately transferred to 25ml of pre-warmed drug free HMI-9 and left to recover for at least 6 hours.

For the parasites transfected with the DNA, the relevant selective drug was added (see Table 2.1) to the flask. The culture was then serially diluted in HMI-9 medium containing the selective drug in 2ml volumes across the wells of a 24-well plate and left for 5-7 days at 37ºC. For the negative control, half of the parasites were removed from the flask and plated in 2ml volumes into the wells of a 24-well plate. This control could confirm that the cells that were transfected did not have underlying growth problems. To the remainder of the cells within the culture flask, the drug was added and the cells serially diluted across the remaining wells of the 24-well plate as described previously. Transfectants were identified by microscopy once the control cells had died, and were grown for further characterisation.

2.7 Fluorescence microscopy analysis of cell cycle status

To detect the cell cycle status of trypanosomes, the cells were incubated with 4',6-diamidino-2-phenylindole (DAPI), which stains the nuclear and kinetoplast DNA. Microscopic analysis of cell cycle status was performed with either *in vivo* derived trypanosomes on methanol-fixed blood smears or paraformaldehyde-fixed *in vitro* derived trypanosomes.

For methanol-fixed blood smears, the slide was removed from methanol, rehydrated in 1X PBS for at least 5 minutes and then stained with ~50µl 10µg/ml DAPI in 1X PBS for 2 minutes in the dark. The slides were washed once in 1X PBS and mounted with mowiol containing DABCO (see below). The slides were dried overnight in the dark and then stored at 4ºC until use.

For the paraformaldehyde fixation of culture derived cells, 1-4x10^6 cells were centrifuged at 1500g, washed twice with 1X PBS and then resuspended in 125µl of cold 1X PBS and 125µl of 8% paraformaldehyde. The cells were incubated on ice for 10 minutes, centrifuged at 1500g for 5 minutes and the paraformaldehyde removed. The cells were resuspended in 130µl of glycine to remove the remainder of the paraformaldehyde and stored at 4ºC for at least an hour. The cells were then centrifuged at 1500g for 5 minutes and the pellet resuspended in 100µl of 1X PBS. Cells could be stored at 4ºC until further use.

An 8 well 6mm microscopy slide (Thermo Scientific) was washed with hot water, ethanol and distilled water and left to dry at room temperature. To each well, 15µl of poly-lysine (Biochrom) was added and the slide incubated in a 'humidity chamber' (a large slide box covered in aluminium foil and filled with wet tissue paper). After 15 minutes the wells were rinsed with dH_2O and dried at 37°C. 10µl of fixed trypanosomes were added to a poly-lysine treated well and incubated for one hour inside the humidity chamber at room temperature. After an hour, the excess 1X PBS was aspirated from the wells and the affixed cells then stained with 10µg/ml DAPI in 1X PBS for 2 minutes at room temperature within the humidity chamber. The wells were washed five times with 1X PBS and once the slides were dry they were mounted with mowiol containing DABCO, as before.

Mowiol + DABCO

10% (w/v) mowiol 4-88 reagent (Calbiochem)

25% (v/v) glycerol

100mM Tris, pH8.5

2.5% (v/v) DABCO

The cells were visualised under Zeiss Imager Z2 microscope and 250 counted per sample. Images were captured using a QImaging Retiga 2000R Camera and edited with ImageJ software (NIH).

2.8 Immunofluorescence analysis

Immunofluorescence analysis (IFA) was used to detect VSG AnTat1.1 or PAD1 expression using specific antibodies. IFA was performed with either *in vivo* derived trypanosomes on methanol-fixed blood smears or paraformaldehyde-fixed *in vitro* derived trypanosomes. For each VSG AnTat1.1 IFA, a positive control (wild type *T. brucei* AnTat1.1 90:13 cells), a negative control (wild type *T. brucei* Lister 427 cells) and a secondary antibody-only control were used. For the PAD1

experiments, the positive controls were *T. brucei* AnTat1.1 90:13 stumpy cells and the for the negative controls, *T. brucei* AnTat1.1 90:13 slender cells. A 'secondary antibody-only' control was also included.

For methanol-fixed blood smears, the slide was removed from methanol, rehydrated in 1X PBS for at least 5 minutes and then blocked for 30 minutes in 2% BSA/1X PBS in a humidity chamber. The slide was washed once in 1X PBS and then incubated for 1-2 hours with the appropriate dilution of primary antibody in a humidity chamber. The primary VSG AnTat1.1 antibody (Jay Bangs, USA) was used at 1:20,000 diluted in 2% BSA/1X PBS and the primary PAD1 antibody (Eurogentec) at 1:1000 diluted in 2% BSA/1X PBS. Excess primary antibody was removed and the slide washed for 3x5 minutes with 1X PBS before incubation with the respective secondary antibodies for one hour in the humidity chamber. For detecting VSG AnTat1.1 and PAD1 expression, an Alexa Fluor® conjugated α-rabbit secondary antibody (Invitrogen) was used at 1:1000 and 1:500 diluted in 2% BSA/1X PBS respectively. The slide was washed three times with 1X PBS and then DAPI stained and mounted as previously described (Section 2.7).

For VSG AnTat1.1 and PAD1 detection from *in vitro* derived cells, the cells were paraformaldehyde fixed, the microscopy slide poly-lysine treated and the cells fixed to the wells as previously described (Section 2.7). The samples were blocked for 45 minutes with 2% BSA/1X PBS in a humidity chamber at 37°C. The excess block was removed from the wells and the respective primary antibody added to the well (antibody concentrations were as used for the methanol fixed cells, see Appendix A). After 45 minutes incubation at 37°C within the humidity chamber, the wells were washed 5 times in 1X PBS and then incubated for a further 45 minutes with the relevant secondary antibody at 37°C within the humidity chamber. The wells were washed a further five times with 1X PBS and then DAPI stained and mounted as previously described.

2.9 Standalone FACS analysis

FACS analysis (out with of the FACS-based VSG switch assays) was used to quantitate VSG AnTat1.1, VSG 221 or GFP expression within a population of cells. For each FACS experiment, a positive control (known VSG AnTat1.1/VSG 221 or GFP positive cells), a negative control (wild type *T. brucei* Lister 427 cells for VSG AnTat1.1 staining and wild type *T. brucei* AnTat1.1 90:13 cells for VSG 221 staining) and a secondary antibody-only control (except for VSG221 staining where the primary antibody was conjugated to a fluorophore) were also analysed to aid gating and confirm correct staining.

For each sample, 2-5x10^6 culture-derived cells were centrifuged at 1500g and washed three times in 1X PBS. If cell viability staining was required, the cells were then incubated in 100µl of the appropriate dilution of Zombie NIR™ dye (Biolegend) for 20 minutes in the dark. *T. brucei* AnTat1.1 90:13 cells were incubated in a 1:20,000 dilution of Zombie in 1X PBS and *T. brucei* Lister 427 cells in a dilution of 1:50,000 in 1X PBS. An unstained sample served as a control. After viability staining, the cells were washed once with 1X PBS to remove excess Zombie stain. Washed cells were resuspended in 500µl of FACS fixing solution (2% formaldehyde, 0.05% glutaraldehyde in 1X PBS). Once in the fix, the cells were incubated overnight at 4^{o}C in the absence of light. After fixation the cells were transferred to 5ml FACS tubes (Corning) and washed three times in 1X PBS. For detection of VSG AnTat1.1 expression, the cells were incubated with 200µl of 1:20,000 diluted α-VSG AnTat1.1 in 2% BSA/1X PBS for 30 minutes. The cells were washed twice more with 1X PBS and then incubated with 200µl of 1:1000 diluted α-rabbit secondary antibody conjugated to Cy5 (Jackson ImmunoResearch) in 2% BSA/1X PBS for a further 30 minutes. The VSG 221 antibody (Monica Mugnier, USA) was conjugated to the Alexa Fluor® 647 fluorophore and therefore only one staining step was required. After the block and wash step, the cells were incubated in 200µl of 1:5000 diluted α-VSG 221/Alexa Fluor® 647 for 30 minutes in the dark. After incubation with the respective fluorophores, the cells were washed three times with 1X PBS and resuspended in 500µl of filtered 1X PBS. The samples were kept in the dark at 4^{o}C until analysis. An antibody was not required to detect GFP fluorescence.

Samples were processed on an LSRII Flow Cytometer (BD Biosciences) and 10,000 events were acquired for each sample. Analysis was performed using the FlowJo 10.4.2 software. Fluorescence compensation was applied to the data analysis when samples had been stained for VSG AnTat1.1 positivity and cell viability using Zombie dye simultaneously. This was because the Cy5 fluorochrome and the fluorescent dye had a small region of spectral overlap. The data was filtered using the forward (FSC) and side scatter (SSC) profiles to ensure that only intact and singular cells were analysed for their fluorescence and the gates for VSG AnTat1.1, VSG 221, GFP or Zombie dye positivity were determined by the controls and kept consistent for each sample within an experimental run.

2.10 RNA extraction

RNA was extracted from 2-5x10^7 cells using the Qiagen RNeasy®Mini Kit as per the manufacturer's instructions. If the extracted RNA was to be used for RT-PCR, qRT-PCR, VSGseq or RNAseq applications, then DNase treatment was performed using the TURBO DNA-*free*™ Kit (Ambion) to remove contaminating gDNA. The manufacturer's protocol was followed using 4U of TURBO™ DNase. RNA content was quantified using a Nanodrop™ Spectrophotometer (Thermo Fisher Scientific). All RNA samples were stored at -80ºC.

If RNA clean-up was required, RNA was precipitated overnight at -80ºC in 0.1 volumes 3M sodium acetate and 2.2 volumes ice-cold 100% ethanol. Samples were centrifuged at 15000g for 30 minutes, then washed in ice-cold 70% ethanol and centrifuged at 15000g for 20 minutes. The pellet was dried and resuspended in RNase-free water (Qiagen).

2.11 RT-PCR

The SuperScript™ III First-Strand Synthesis System (Invitrogen) was used for first-strand cDNA synthesis. For each sample, 2 reactions were performed in parallel- one reaction containing reverse transcriptase (+RT) and one without (-RT)

to confirm that the DNase treatment performed after the RNA extraction step was sufficient. For VSG-specific RT-PCR reactions, 2pmol of the '14MER_RVS' primer (Table 2.2) was used for the first-strand synthesis reaction. For all other RT-PCR reactions, 50µM of the oligo$(dT)_{20}$ primer was used. To each 0.5ml reaction tube, the following components were added:

200ng-1µg RNA

1µl 10mM dNTP mix

1µl primer

RNase free water to a total volume of 13µl

The tube was incubated at 65ºC for 5 minutes, before being placed on ice for at least one minute. After cooling, the contents of the tube were transferred to a PCR tube containing the following:

4µl 5X first-strand buffer

1µl 0.1M DTT

1µl RNaseOUT™

1µl 200U/µl SuperScript III RT (for -RT samples this was substituted for 1µl dH_2O)

In a PCR block, the reaction was incubated for 60 minutes at 50ºC. The reaction was then inactivated by incubation at 70ºC for 15 minutes. To remove RNA complementary to the cDNA, 1µl (2U) of *E. coli* RNase H was added to each sample and the tubes incubated at 37ºC for 20 minutes. Amplification of double-stranded cDNA was performed by conventional PCR (see Section 2.5.2) using gene-specific primers. 1µl of a 1:50 dilution of the single stranded cDNA was used for each reaction.

2.12 Quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) reactions were performed on a LightCycler 96 machine (Roche). For each qRT-PCR reaction, single-stranded cDNA, as produced in Section 2.10, was used at a concentration of 1:20. Each 25µl reaction was set up in triplicate with a corresponding water and -RT control. Transcript abundance was quantified relative to the internal control *Tb*ZFP3, for which the reaction conditions differed. Reactions were set up as below:

Samples for testing

12.5µl Power SYBR Green PCR Master Mix (Applied Biosystems)

1.5µl 10µM FWD primer

1.5µl 10µM RVS primer

7µl dH_2O

2.5µl 1:20 diluted single-stranded cDNA

*Tb*ZFP3 internal control

12.5µl Power SYBR Green PCR Master Mix (Applied Biosystems)

2.25µl 10µM FWD primer

2.25µl 10µM RVS primer

1µl dH_2O

7µl 1:20 diluted single-stranded cDNA

The reaction conditions were as follows (a melt curve was added to the end of the programme to ensure amplification to a single product):

10 minutes at 95°C

40 x (15 seconds at 95°C, 1 minute at 60°C)

Melt curve of 60-95°C at a rate of 0.03°C/second

Transcript abundances were quantified relative to ZFP3 using the ΔΔCT method and then compared.

2.13 Northern Blot

Prior to performing the blot, all the required equipment was first washed with detergent, rinsed with dH_2O, sprayed with 70% ethanol, and then left to air dry. The DIG-labelled *HYP2*, *NEK* and *DYRK* RNA probes used in this study were generated by Dr Eleanor Silvester, Dr Lindsay McDonald and Dr Paula MacGregor, respectively, using the Roche DIG RNA Labelling Kit.

RNA was prepared as Section 2.9 (for the purposes of a Northern blot, DNase treatment was not necessary). For each sample, 1µg of RNA was added to 9µl of formamide, 3µl of formaldehyde, 2µl of 10X MOPS and 2µl of RNA loading buffer (see below). The samples were incubated on a heat block at 65^{o}C before being loaded onto a 1.2% RNA gel (see below). The gel was electrophoresed in 1XMOPS running buffer for 90 minutes at 150V.

RNA loading buffer	RNA gel
30% Formamide	1.2 % Molecular grade agarose (Bioline)
16.6% Formaldehyde	1X MOPS
1X MOPS	3% Formaldehyde
0.01% Bromophenol blue	
10% Glycerol	

The RNA gel was stained with 1µg/ml ethidium bromide in 1X MOPS for 15 minutes before being destained twice with dH_2O for at least 30-60 minutes. The RNA was visualised and photographed on a Syngene G:box UV transilluminator.

RNA was transferred overnight to a positively charged nylon membrane (Roche) by capillary transfer. This transfer step was set up as shown in Figure 2.1.

Briefly, a large plastic box was filled to a level of approximately 3 cm with 10X SSC (1.5M NaCl, 0.15M tris sodium citrate). Over this, a flat piece of Perspex was placed such that the sides of the box were covered, but a small gap remained at the top and bottom of the tray. A sheet of filter paper pre-soaked in 10X SSC was placed over the Perspex with its ends positioned through the gaps and into the 10X SSC in the box below. The RNA gel was then placed upside down on top of the filter paper, and clingfilm used to cover the edges of the box to prevent evaporation of the 10X SSC. The positively charged membrane, pre-soaked in 10X SSC was placed on top of the RNA gel, followed by two sheets of filter paper soaked in 2X SSC, a block of paper towels and a large catalogue to act as a weight.

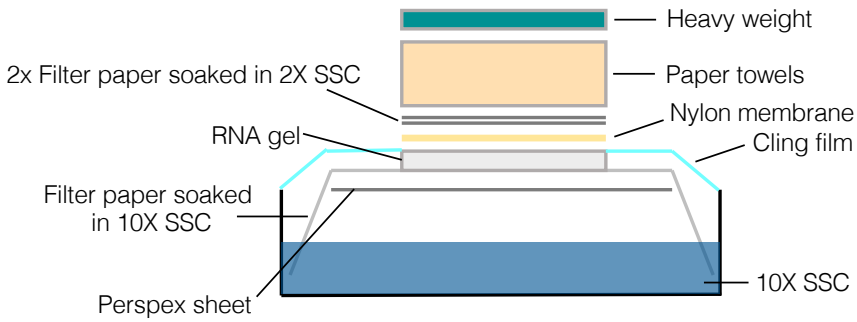


Figure 2.1 Northern Blot Transfer Setup

The next day, the transfer was disassembled and the RNA gel visualised under a UV transilluminator in order to confirm that no RNA remained on it. The membrane was left to dry at room temperature. A Stratalinker (Uvitec) at 0.12 joules was used to UV cross-link the RNA to the membrane.

The membrane was prehybridised for one hour at 68ºC in 10ml of Hybridisation buffer (see below) within a Techne Hybridiser HB-1D hybridisation oven. After 1 hour incubation, the buffer was removed and 600ng of the appropriate RNA probe (which had been boiled in 100µl of hybridisation buffer for 7 minutes just prior) then added to the membrane along with 7ml of pre-warmed hybridisation

buffer. The membrane was hybridised overnight at the appropriate temperature. For the *HYP2* and *NEK* probes this was 68ºC, and for the *DYRK* probe this was 62ºC.

Hybridisation Buffer

5X SSC

50% Formamide

0.02% SDS

2% DIG Block

After the overnight hybridisation, the probe was removed and the membrane washed twice for 30 minutes at hybridisation temperature in 2X SSC/0.1% SDS. The membrane was rinsed once more for 30 minutes at hybridisation temperature in 0.5X SSC/0.1% SDS. In freshly cleaned plastic boxes, the membrane was rinsed for 1 minute in wash buffer (see below) and then incubated for 1 hour on a shaker in 50ml of Maleic acid buffer (see below) with 1% DIG block. After an hour, 2µl of Anti-Digoxigenin-AP Fab Fragments (Roche) were added to the blocking solution and the membrane incubated for a further 30 minutes with agitation. The membrane was then washed for 3x10minutes in wash buffer.

Maleic Acid Buffer	Wash Buffer
100mM Maleic acid	99.7% Maleic acid buffer
150mM NaCl	0.3% Tween 20
pH7.5, autoclave	

The membrane was soaked evenly in Detection buffer (see below) for 2 minutes on a shaker and then placed between two sheets of polyethylene. To the membrane, 250µM of CDP-Star chemiluminescent substrate (Roche) in detection buffer was added and then left for 2 minutes. After two minutes, the excess CDP

Star/detection buffer was removed and the polyethylene sheets sealed together. The membrane was incubated in the dark at 37°C for 15 minutes and then placed within an X-ray cassette and exposed to Super RX Medical X-ray film (Fuji) within a dark room. Initially films were exposed for 10 minutes, with further exposure times being determined from the result of this exposure. Films were developed on a MI-5 X-ray film processor.

2.14 Western Blot

Western blotting was used to detect the expression of the stumpy-specific marker, PAD1, in cells isolated on the final day of an *in vivo* infection. Once the trypanosomes had been purified from blood (see Section 2.4), $4x10^7$ cells were centrifuged at 1500g for 5 minutes, washed twice in 1X PBS and then resuspended in 100µl of Laemmli Buffer (see below). Samples were stored at -80°C until further use.

Laemmli Buffer

62.5mM Tris, pH6.8

2% SDS

10% Glycerol

5% 2-mercaptoethanol

0.0025% bromophenol blue

PAD1 samples were not boiled prior to resolving on a sodium dodecyl sulphate (SDS) gel, but alternatively sonicated for 2x30 seconds using a Bandelin Sonorex sonicator. Samples were cooled on ice between sonications. For identifying PAD1, a 55kDa protein, 10% resolving gels and 4% stacking gels were prepared (see below). The resolving gel was poured between two glass plates (a 0.75 or 1.5mm spacer plate and a short plate) that were clamped together using a casting frame and stand (Biorad) until it reached a level approximately 2cm below the top of the short

plate. The gel was covered in a thin layer of isopropanol and left to set at room temperature. Once the resolving gel had set, the isopropanol was removed, the gel rinsed with dH_2O, and the stacking gel added until it reached the top of the short plate. Depending on the number of samples to be resolved, either a 10 or 15-well comb was inserted between the two plates, and the gel left to set. Once set, the comb was removed, the gel was inserted into the assembled gel running apparatus and the tank filled 1X TGS Running Buffer (see below).

10% Resolving Gel	4% Stacking Gel
375mM Tris, pH8.8	125mM Tris, pH 6.8
0.1% SDS	0.1% SDS
10% Acrylamide	4% Acrylamide
0.4% APS	0.1% Ammonium persulphate
0.05% TEMED	0.1% TEMED

APS and TEMED were added last.

1X TGS Running Buffer

25mM Tris

192mM Glycine

0.1% SDS

For each sample, $4x10^6$ cells/sample were loaded into a well and electrophoresed at 140V alongside a Novex Sharp Pre-stained Protein Standard (Life Technologies) until the bromophenol blue dye reached the bottom of the gel. Protein was transferred from the SDS gel to a nitrocellulose membrane (GE Healthcare) using a Bio-Rad Mini Trans-Blot® Cell that was assembled according to the manufacturer's instructions. The tank was filled with 1X Transfer Buffer (see below) and the transfer carried out at 200mA for 90 minutes.

1X Transfer Buffer

25mM Tris

187mM Glycine

0.1% SDS

20% Methanol

After the transfer, the membrane was removed from the apparatus and stained with Ponceau Red (Sigma-Aldrich) for 5 minutes to confirm that the protein had been transferred. The Ponceau stain was removed by incubation with 1X PBS/0.05% Tween (PBST) for 1 hour at room temperature with agitation. The membrane was blocked for 1 hour at room temperature, or overnight at 4ºC, with 5% milk in PBST and agitation and after this incubated for a further hour at room temperature, or overnight at 4ºC, with 1:1000 PAD1 1º antibody diluted in PBST. The membrane was then washed for 3x5 minutes in PBST.

For secondary detection of the PAD1 antibody, the LI-COR Odyssey system was used. The membrane was incubated in the dark for 1 hour at room temperature with an α-rabbit antibody conjugated to the IRDye® 800CW fluorophore (LI-COR) diluted 1:7000 in 50% LI-COR Blocking Solution and 50% PBST. The membrane was then washed twice for 5 minutes in PBST and once more in 1X PBS only, with agitation, before being scanned using the LI-COR Odyssey Imager.

EF1α was used as a loading control for the Western blot. Detection of EF1α was performed as described for PAD1, however using the anti-EF1α antibody (Merck Millipore) at a concentration of 1:7000 and an α-mouse/IRDye® 560CW secondary fluorophore conjugated antibody at 1:5000.

2.15 Southern Blot

The DIG-labelled *PAC* DNA probe was generated from 1µg of PCR-amplified *PAC* DNA using the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche) according to the manufacturer's instructions.

For each replicate sample, 1µg of gDNA was digested overnight with the appropriate restriction enzyme (in a total volume of 20µl) at 37ºC. The next day the samples were loaded into the wells of a 0.8% agarose gel, alongside a DNA ladder, and electrophoresed at 120V until they had run approximately two-thirds of the way down the gel's length. The fragments were visualised under a UV transilluminator and a photo taken with a ruler aligned to the bottom of the wells and parallel to the DNA ladder (this acted as a reference for band size once the membrane was probed and detected). The gel was then soaked in at least 300ml of Denaturation Solution (see below) for 2x15mins with agitation, washed with dH_2O, and then soaked for a further 2x15 minutes in Neutralisation Buffer (see below).

Denaturation Solution	Neutralisation Solution
1.5M NaCl	1M Tris
0.5M NaOH	1.5M NaCl
	pH 7.4

Transfer of the DNA to a positively-charged membrane, and the subsequent UV cross-linking of the DNA to the membrane, was set up exactly as described for Northern blotting in Section 2.12.

The membrane was pre-hybridised with 10ml/100cm^2 membrane of pre-heated DIG EASY HYB Granules (Bottle 7 of DIG High Prime DNA Labelling and Detection Starter Kit II, Roche) for at least 30 minutes at 55ºC. During the pre-hybridisation stage, 25ng/µl *PAC* probe in a volume of 50µl was denatured by boiling at 95ºC for 5 minutes and then rapidly chilled on ice. After 30 minutes the hybridisation solution was poured off and replaced with 7ml fresh hybridisation solution containing the denatured probe. The membrane was hybridised overnight in a hybridisation oven at 55ºC. The membrane was washed twice for 5 minutes at room temperature in 2X SSC/0.1% SDS with agitation and then returned to the hybridisation tube and incubated for 2x15 minutes at 68ºC in 0.5X SSC/0.1% SDS.

Immunological detection of the membrane-bound probe was performed exactly as described for Northern blotting in Section 2.12.

2.16 FACS-based VSG switch assays

2.16.1 Growth phase

For each experiment, a 30ml logarithmic culture was grown to ~1x10^6 cells and 2x10^7 cells were removed and fixed for FACS analysis as described in Section 2.9. From the remaining culture, 1ml was diluted to 6 cells/ml in a volume of 10ml (for one experimental condition only) or 20ml (for –/+ DOX comparisons). 10ml was spread over a single 96-well plate in 100µl volumes using a multichannel pipette. For -/+ DOX comparisons, 10ml of uninduced cells were plated first and 1µg/ml doxycycline then added to the remaining 10ml of cells before they were plated out over another 96-well plate. The plates were incubated at 37^{o}C and scored for positive wells by microscopy after 3-6 days. Typically, each plate returned 5-12 positive wells.

Once the populations were judged by eye to be dense enough for dilution (but not exceeding 1x10^6 cells/ml), three wells were selected to be 'scaled-up' in parallel in increasing volumes of HMI-9 until there were at least 2x10^7 cells. For +DOX samples, doxycycline was added to the fresh HMI-9 upon culture dilution. The cells were washed and fixed for FACS analysis as described in Section 2.9.

2.16.2 FACS analysis

Flow cytometry was used to quantitate VSG AnTat1.1 or VSG 221 and GFP expression within a population of cells. For each FACS experiment staining controls, as described in Section 2.9, were used. For the positive control, 2x10^7 cells were fixed and processed for analysis, whereas 2-5x10^6 cells were fixed and processed for the negative and 'secondary only' controls. Each sample was processed for flow cytometry as described in Section 2.9, except that the blocking and antibody steps were performed in volumes of 1ml. The labelled cells were resuspended in 1ml of filtered 1X PBS and to each experimental tube, 40µl of vortexed CountBright

Absolute Counting Beads (Thermo Fisher Scientific) were added. The number of beads within this 40µl varied depending on Lot number, but was typically ~$5x10^4$. To ensure statistical significance, at least 1,000 bead events were collected per experimental sample. Samples were processed on an LSRII Flow Cytometer (BD Biosciences) and 1,000,000 events were acquired for each experimental sample and the positive control. 10,000 events were collected for the negative and secondary only controls.

2.16.3 Data interpretation

Analysis was performed using the FlowJo 10.4.2 software. The data was filtered using the forward (FSC) and side scatter (SSC) profiles to ensure that only intact and singular cells were analysed for their fluorescence.

The first gate to be applied to the population was that determining VSG AnTat1.1 or VSG 221 positivity. To aid gating, the relevant negative, secondary antibody-only (for VSG AnTat1.1 staining only) and 'Day 0' positive controls were used. The VSG AnTat1.1/221 gates were placed directly to the left of the positive control peak to account for cells taking up to 4.5 days to fully replace their VSG surface coat following a genetic switch. The cells gated as VSG AnTat1.1 or VSG 221 negative were next analysed for GFP positivity. A similar gating strategy to that used to define VSG AnTat1.1 or VSG 221 positivity was used to interrogate GFP positivity since, like a VSG coat taking time to be fully replaced, GFP fluorescence would not immediately dissipate following an *in situ* switch of VSG expression. Therefore, the gate defining GFP positivity or negativity was placed directly to the left of the 'Day 0' positive control peak. Gating was kept consistent between samples of the same experiment.

To calculate the number of VSG switches/cell from the FACS data, the method described by Schulz et al. (2016) was followed. To convert this value to VSG switches/cell/generation, the number of VSG switches per cell was divided by the number of population doublings that were calculated in Section 2.16.1 The number of VSG AnTat1.1 or VSG 221 negative and GFP positive cells was divided by the total number of VSG AnTat1.1 or VSG 221 negative cells to calculate the proportion

of cells that had switched VSG expression by DNA recombination downstream of the GFP reporter construct.

2.17 VSGseq workflow

VSGseq analysis was performed in collaboration with Dr Monica Mugnier (Johns Hopkins Bloomberg School of Public Health, USA). Parasite growth, magnetic activated cell sorting (MACS) and RNA extraction was performed at the University of Edinburgh, whilst the RT-PCR, library preparation and sequencing were performed at the Johns Hopkins Bloomberg School of Public Health. The pipeline used for data analysis (Section 2.17.5) was developed by Dr Monica Mugnier.

2.17.1 Growth phase

Triplicate populations of -/+ DOX GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi cells were expanded from a single cell as described in Section 2.17. However, since a greater number of cells were required, one further 'scale-up' step was performed, bringing the final culture volume to 125ml.

2.17.2 Magnetic-activated cell sorting

Magnetic-activated cell sorting (MACS) was performed to enrich populations of GFP^{ESpro}ANTat1.1^{ES} cells for cells that did not express VSG AnTat1.1. The MACS protocol for the depletion of VSG2 from an *in vitro* culture of *T. brucei* Lister 427 cells (Schulz et al., 2016) was followed with the following alterations: an α-VSG AnTat1.1 antibody was used at a concentration of 1:50; cells were resuspended in 180µl of the primary antibody dilution instead of 150µl; cells were rotated on an orbital wheel at 4^{o}C and not vortexed; 120µl of anti-rabbit magnetic-activated cell sorting microbeads (Miltenyi Biotec) were used instead of 110µl; and, where the number of cells after pre-MACS RNA, IFA and FACS sample extraction was >1x10^{8} cells, samples were equally split in two and loaded onto two separate columns (to prevent overloading of the column and contamination of the flow-through with VSG AnTat1.1 positive cells) with the final flow-throughs being pooled. From the column

flow-through, 1ml was washed and fixed for IFA analysis of VSG expression (Section 2.8) and the remainder used for the post-MACS RNA extraction (Section 2.17.3).

2.17.3 RNA preparation and glycogen precipitation

RNA was extracted from 25ml of each of the pre-MACS populations and from the entirety of the MACS flow-throughs (after the removal of 1ml cells for IFA, Section 2.17.2) as described in Section 2.10. The final elution was performed with 30µl of RNase-free dH_2O.

To purify and concentrate the pre-MACS RNA samples, the RNA was ethanol precipitated and resuspended in 10µl RNase-free dH_2O as described in Section 2.10. To aid in the recovery of the low concentrations of nucleic acids present in the post-MACS RNA samples, each of the post-MACS RNA samples were co-precipitated with the nucleic acid carrier glycogen (Thermo Scientific) in the following reaction:

30µl RNA

3µl 3M Sodium acetate

0.75µl 20mg/ml glycogen (Thermo Scientific)

75µl 100% ice-cold ethanol.

The samples were incubated overnight at -80^{o}C and the RNA then pelleted and resuspended in 10µl RNase-free dH_2O as before.

2.17.4 RT-PCR and library preparation

First strand cDNA was produced using the SuperScript™ III First-Strand Synthesis System (Invitrogen), but with different reaction conditions to those detailed in Section 2.11. No 'without reverse transcriptase' reaction was performed. For each sample, the following reaction was set up in a 0.2ml PCR tube:

8µl of RNA

1µl of 2mM VSG 3' UTR primer

1µl dNTP mix

The samples were incubated at 65ºC for 5 minutes and then cooled on ice for at least 1 minute. After cooling, the following synthesis mix was added to each reaction tube:

2µl RT Buffer

4µl 25mM $MgCl_2$

2µl 0.1M DTT

1µl 40U/µl RNaseOUT

1µl 200U/µl Superscript III RT

The samples were incubated for 10 minutes at 25ºC, followed by 50 minutes at 50ºC. The reaction was terminated by incubation at 85ºC for 5 minutes. To each sample, 1µl of RNase H and 1µl of RNase A (Qiagen) were added, followed by 30 minutes incubation at 37ºC. Reaction purification was performed using the AMPure RNAclean XP Kit (Beckman Coulter) according to the manufacturer's instructions. The final elution was performed with 32µl of dH_2O. To synthesise double-stranded cDNA from the first strand synthesis products, the following reaction was set up for each sample in 0.2ml PCR tubes:

32µl single strand cDNA

10µl Phusion HF buffer

1µl 10mM dNTPs

2.5µl 10µM 14mer-SP6 oligo

2.5µl 10µM SL oligo

1.5µl DMSO

0.5µl Phusion DNA Polymerase

The samples were put into a thermocycler with the following programme:

98°C	30 seconds	
98°C	10 seconds	x22 cycles
55°C	10 seconds	
72°C	45 seconds	
72°C	5 minutes	

The PCR reactions were purified using the AMPure XP Kit (Beckman Coulter) according to the manufacturer's instructions, with the final elution being performed in 30µl dH_2O. DNA was quantified using the Qubit HS dsDNA assay kit (Thermo Fisher) according to the manufacturer's instructions.

For library preparation, the NextEra XT Library Preparation system (Illumina) was used according to the manufacturer's protocol. The NextEra Library Preparation Kit uses engineered transposomes to tagment double-stranded gDNA. Tagmentation is a process whereby gDNA is fragmented and tagged with adapter sequences simultaneously. The limited-cycle PCR uses the adapter sequences to amplify the DNA, whilst also adding the index adapter sequences onto both of the DNA that allow for the dual-indexed sequencing of pooled libraries on Illumina sequencing platforms (Illumina, 2018). The PCR reaction was cleaned up using the AMPure XP system as before, with the final elution being performed in 27µl of resuspension buffer.

The libraries were pooled such that 0.02pmol of each library was present in a mixture volume of 103.5µl. The pooled libraries were loaded onto a single Illumina Flow Cell and 100bp single-end sequencing was performed in-house on an Illumina HiSeq 2500.

2.17.5 VSGseq data processing

Processing of the VSGseq data through the pipeline provided at https://github.com/mugnierlab/VSGSeqPipeline/blob/master/VSGSeqPipeline.py,

was kindly performed by Dr Monica Mugnier. Adapters (from the tagmentation step) and SP6 sequences (from the *VSG* PCR step) were trimmed using trim_galore (Babraham Bioinformatics) and cutadapt (Martin, 2011), respectively. Sequencing reads were assembled by Trinity (Grabherr et al., 2011) and ORFs (defined as a start codon to stop codon or a start codon to the end of a contig) of >900bp identified. These ORFs were compared against the *T. b. brucei* EATRO1125_vsgs reference database (http://129.85.245.250/Downloads/vsgs_tb1125_all_atleast150aas_cds.txt) using BLASTn to identify *VSG* sequences. Sequences that matched chromosomal sequences or non-*VSG* sequences were removed.

Since the Bowtie alignment programme (used later in the pipeline) cannot distinguish between *VSG* sequences with >98% sequence similarity using this pipeline's parameters, *VSG*s with >98% sequence identity were merged using cd-hit-est (Weizhong and Godzik, 2006). These final merged contigs represented an individual VSG, which was included in the reference genome. The reads for each sample were aligned to the reference genome using Bowtie, allowing for only uniquely mapping reads and no more than 2 mismatches per read. The output MULTo analysed csv. file showed the expression of each VSG in each sample, both in terms of RPKM (reads per kilobase of transcript, per million mapped reads) and percentage of the population (RPKM for that VSG/total RPKM), and the BLAST similarity results for each de novo assembled *VSG* compared to the most similar reference *VSG*.

2.18 RNAseq

For the RNAseq analysis, RNA was prepared from triplicate populations of the parental 'Start' GFP^{ESpro}AnTat1.1^{ESpro} *NEK* RNAi cells, the monomorphic 'End' GFP^{ESpro}AnTat1.1^{ESpro} *NEK* RNAi cells and two intermediate samples from the passage time-course, named INT1 and INT2 respectively.

The cells were seeded at 2x10^3 cells/ml in 50ml HMI-9 medium (except for the selected monomorphs which were seeded at 1x10^3 cells/ml due to their shorter doubling time) and incubated at 37^{o}C for 48 hours. The density of the cultures never

exceeded $6.8x10^5$ cells/ml and therefore it is highly unlikely that any differences in gene expression between the pleomorphic and monomorphic cells are due to a density-dependant response to the accumulation of SIF in the pleomorphic populations. After 48 hours, the contents of the flasks were centrifuged and RNA prepared and TURBO™ DNase treated as described in Section 2.10. RNA quality was checked by electrophoresing 200ng of RNA for each sample on an RNA formaldehyde gel (Section 2.13) and RNA quantity was determined by a Nanodrop™ Spectrophotometer (Thermo Fisher Scientific). Samples were stored at -80°C until shipment.

For each replicate, at least 1.3µg of RNA was sent to BGI Tech Solutions (Hong Kong) for sequencing using the 'Eukaryotic Transcriptome Resequencing HiSeq 4000 PE100' service. 4Gb of clean data was generated for each RNA sample. Raw RNA sequencing data was kindly processed by Dr Alasdair Ivens. Quality control of the raw data was performed using the Fast QC programme (Babraham Bioinformatics), and paired end reads were trimmed using cutadapt (Martin, 2011). Reads were aligned to the *T. b. brucei* TREU927 genome using Bowtie 2 (Langmead and Salzberg, 2012) to obtain counts for the number of reads that mapped to each gene. These counts were normalised to the reads per kilobase of transcript per million mapped reads (RPKM) to account for gene size and read depth. Group-wise comparisons between the 'Start', 'INT1', 'INT2' and 'End' datasets were performed using the software package limma (Ritchie et al., 2015).

2.19 Identification of single nucleotide polymorphisms

For the identification of single nucleotide polymorphisms (SNPs) DNA was prepared from a single parental 'Start' $GFP^{ESpro}AnTat1.1^{ESpro}$ *NEK* RNAi population and a the monomorphic 'End' $GFP^{ESpro}AnTat1.1^{ESpro}$ *NEK* RNAi population.

The parental pleomorphic population was seeded at $2x10^3$ cells/ml in 100ml HMI-9 medium and the selected monomorph population at $1x10^3$cells/ml in 100ml HMI-9 medium. The cells were cultured for 48 hours at 37°C before the contents of the flask were centrifuged and DNA prepared from the pellet as described in Section

2.5.1. DNA quality was assessed by electrophoresing 500ng of DNA for each sample on an agarose gel (Section 2.5.3) and DNA quantity was determined by a Nanodrop™ Spectrophotometer (Thermo Fisher Scientific).

For each replicate, >2µg of DNA was sent to BGI Tech Solutions (Hong Kong) for whole genome sequencing (WGS) using the 'Illumina plant and animal whole genome resequencing HiSeq 4000' service. At least 30X genome coverage was generated for each DNA sample.

The processing of raw sequencing data, identification of SNPs/Indels and mapping of SNP regions to the reference *T. b. brucei* TREU927 genome were kindly performed by Dr Alasdair Ivens. Reference 'Start' and 'End' genome assemblies were made using CLC Assembly Cell (Qiagen). The 'Start' and 'End' reads were mapped to the reference 'Start' genome using Bowtie 2 (Langmead and Salzberg, 2012) and SNPs/indels called using SAMtools (Li et al., 2009). A Variant Call Format (VCF) (Danecek et al., 2011) file was generated from the SNP/indel scan output and a QUAL score cut-off (the Phred-scaled probability that a polymorphism exists at a given site) of 200 applied to the data. For each SNP, a 301bp region (the SNP and 150 base pairs to either side) were mapped to the *T. b. brucei* TREU927 reference genome so that they could be viewed in the context of annotation.

2.20 Statistical analysis

All statistical analysis performed in this thesis used the GraphPad Prism 7.0d software. P values ≤ 0.05 were considered statistically significant.

3 : COMPARING THE VSG SWITCH RATE OF PLEOMORPHIC AND INDUCIBLY MONOMORPHIC POPULATIONS *IN VITRO* BY FLOW CYTOMETRY

3.1 Introduction

Studies of trypanosome antigenic variation preferentially investigate laboratory-adapted cell lines which are antigenically homogenous and exhibit high levels of antigen stability. These trypanosomes have been cyclically passaged through laboratory rodents, thus selecting against tsetse transmission potential and generating highly virulent populations of monomorphic trypanosomes. In the relatively few studies of antigen switch frequency in recently isolated pleomorphic parasites, it has been estimated that the transmission-competent cells switch their expressed VSG at a rate 66,000-fold greater than laboratory adapted trypanosomes (Turner and Barry, 1989). It is not yet known if this loss of stumpy differentiation capacity is directly linked to reduction in VSG switch frequency, or whether the two processes, though both selected by multiple passage, are in fact independent.

Mony et al. (2014) identified components of the stumpy formation signalling pathway through a genome-wide RNAi screen, and demonstrated that their inducible gene silencing by RNAi caused pleomorphic parasites to remain slender *in vivo*, despite the accumulation of SIF. These RNAi cell lines therefore provide us with a tool to switch pleomorphism 'on' or 'off' inducibly, without long term passage. By exploiting this inducible monomorphism model, we can interrogate antigen switch frequency in the same cell lines, but where the populations are either differentiation competent or incompetent. By these means we can establish whether antigen switch frequency is intimately linked to the developmental quorum-sensing interactions of the parasite in the mammalian bloodstream.

3.2 Inducible knockdown of *HYP2* and *DYRK* generated inducibly monomorphic trypanosome populations

Pleomorphic *T. brucei* RNAi cell lines were studied, whereupon induction of RNAi against a particular gene involved in the stumpy formation signalling pathway 'turned off' the capacity for the production of stumpy forms (Mony et al., 2014). *Hypothetical Protein 2* (*HYP2*, Tb927.9.4080) and *DYRK* (Tb927.10.15020) are two

genes identified by Mony et al. (2014) as conferring insensitivity to SIF when knocked down in pleomorphic parasites *in vivo*. HYP2 is a potential effector molecule within the signalling pathway. Whilst the protein shares no obvious homology with known genes, a DksA zinc finger motif, involved in the regulation of bacterial ribosomal RNA transcriptional responses to oxidative and nitrosative stress and the post transcriptional control of quorum sensing dependent genes, is present (Henard et al., 2014; Jude et al., 2003; Mony et al., 2014). DYRK is a dual-specificity tyrosine-phosphorylation-regulated kinase which likely functions as a signal transducer upstream of the effector molecule HYP2. The DYRK family of kinases are an evolutionarily conserved family in which many members function in signalling pathways integral to cellular development and homeostasis (Aranda et al., 2011). *Saccharomyces cerevisiae* YAK1, the founding member of the DYRK family, has been shown to be a negative regulator of cell proliferation and overexpression of *Dictyostelium discoideum* YAKA causes cell-cycle arrest (Garrett et al., 1991; Soppa and Becker, 2015). McDonald et al. (2018) recently used an extragenic suppression approach in pleomorphic trypanosomes to demonstrate that NEK17 and protein phosphatase 1 (PP1) driven stumpy formation was DYRK dependent, whilst RBP7 driven stumpy formation was DYRK independent (Figure 1.11).

To generate *HYP2* and *DYRK* RNAi cell lines, Mony et al. (2014) designed RNAi target fragments using the RNAit software (Redmond et al., 2003), amplified the sequences from *T. brucei* AnTat1.1 90:13 gDNA and then cloned the fragments into the pALC14 stem-loop plasmid (Bochud-Allemann and Schneider, 2002; Mony et al., 2014). The cloned plasmids were then transfected into *T. brucei* AnTat1.1 90:13 cells by electroporation. These lines were recovered from storage at -80^{o}C and used for this study.

3.2.1 Knockdown of *HYP2* or *DYRK* in pleomorphic cells conferred partial resistance to a cAMP analogue

To validate the inducibility of the RNAi response, and the concurrent loss of stumpy formation capacity upon induction, logarithmic populations of *HYP2* RNAi and *DYRK* RNAi cells were grown in triplicate either in the absence or presence of 1µg/ml doxycycline *in vitro* (± induction). After 24 hours, the cells were treated with

100µM of the SIF mimic, 8-pCPTcAMP. Incubation with 8-pCPTcAMP causes rapid G1 arrest and limited expression of stumpy-specific genes in both pleomorphic and monomorphic trypanosome strains (Laxman et al., 2006). Upon the induction of *HYP2* or *DYRK* RNAi, cells would be expected to show reduced responsiveness to 8-pCPTcAMP, since these genes form parts of the stumpy formation signalling pathway. No growth effect was observed in either of the two RNAi lines upon incubation with doxycycline alone (Figure 3.1). Uninduced cells grown in the presence of 8-pCPTcAMP showed significant growth inhibition following 24hours exposure, thus demonstrating that the RNAi remained regulatable, but this growth inhibition was only partially relieved upon induction of RNAi against *HYP2* or *DYRK*. The induced *HYP2* RNAi population remained proliferative when grown in the presence of 8-pCPTcAMP, albeit replicating at a slower rate than the untreated population. In contrast, growth in the induced *DYRK* RNAi population stopped after 48 hours incubation with the SIF mimic indicating limited resistance to 8-pCPTcAMP upon knockdown of the transcript.

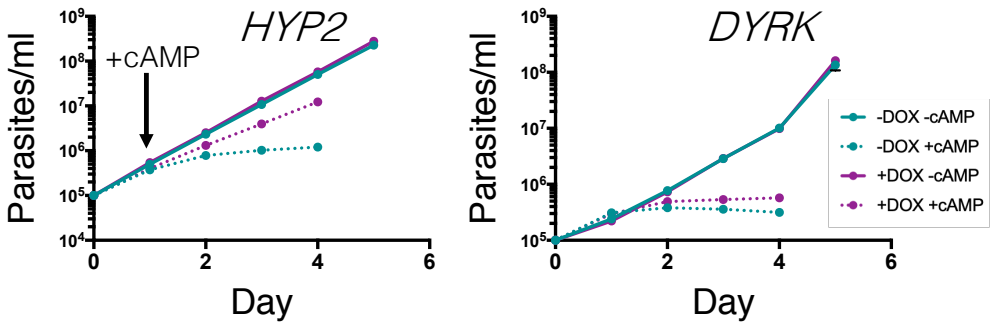


Figure 3.1 Induction of *HYP2* and *DYRK* RNAi *in vitro* confers partial 8-pCPTcAMP resistance. *In vitro* growth of *HYP2* (left) and *DYRK* (right) RNAi cell lines in the presence/absence of doxycycline (-/+DOX) and 8-pCPTcAMP (-/+ cAMP). 8-pCPTcAMP was added 24 hours post doxycycline addition. Induced *HYP2* RNAi cells were partially resistant to 8-pCPTcAMP mediated cell cycle arrest, whereas the *DYRK* RNAi cells only showed very low levels of resistance to the compound. Data represents the mean±S.D (n=3).

3.2.2 Induction of *HYP2* or *DYRK* RNAi in pleomorphic cells generated differentiation defective cells *in vivo*

Since the induced RNAi lines still retained some responsiveness to 8-pCPTcAMP *in vitro*, the cells were tested *in vivo* for their SIF-mediated differentiation competence upon knockdown of *HYP2* or *DYRK*. Triplicate untreated and doxycycline-treated female MF1 mice were infected with either *HYP2* or *DYRK* RNAi cells and observed for their parasitaemia. An untreated 'control' mouse was infected with the parental AnTat1.1 90:13 cells. Blood samples were taken for estimation of the parasitaemia by the rapid 'matching' method (Herbert and Lumsden, 1976) from day 3 onwards. The uninduced *HYP2* RNAi cells restricted their parasitaemia from ~ day 4 onwards (Figure 3.2a) and transformed morphologically to stumpy forms, characterised by their undulating membrane, attached flagellum and stumpy appearance (Figure 3.2b, left panel). Induced *HYP2* RNAi cells also showed a plateau in growth, however at approximately 2-fold greater parasitaemia than the uninduced cells. On the final day of infection, the induced population consisted of morphologically slender, intermediate and stumpy cells (Figure 3.2b, right panel). Consistent with the presence of stumpy cells, Western blotting detected PAD1 expression in protein isolated from cells purified from the blood on day 6 of infection, though at a slightly lower level in the induced cells relative to the uninduced cells (Figure 3.2c). Blood smears were fixed in methanol from day 3 onwards, and used to assess the cell cycle status of 250 cells. Cells in G1 adopt a 1K1N (1 kinetoplast, 1 nucleus) configuration. As cells progress through the cell cycle however, they first replicate their kinetoplast, becoming 2K1N, followed by their nucleus, now 2K2N, before finally dividing. Both groups of cells accumulated in 1K1N as the infection progressed, though this was more rapid in the uninduced cells (Figure 3.2d). Nonetheless, the failure of the induced cells to fully arrest in G0/G1, their diminished expression of PAD1 compared to the uninduced cells and their slender morphology despite being at greater parasitaemia than the uninduced cells, together suggests that the *HYP2* RNAi cells have reduced their ability to respond to the accumulation of SIF.

The differentiation phenotype observed in the *DYRK* RNAi line upon induction *in vivo* was much stronger. Whilst the uninduced cells transformed morphologically to stumpy forms from day 4 onwards (Figure 3.2f, left panel), the induced cells remained proliferative, with the parasitaemia continuing to rise until the experiment had to be ended on humane grounds on day 5 (Figure 3.2e). The induced cells retained the typical long slender morphology and free flagellum associated with proliferative bloodstream forms (Figure 3.2f, right panel). PAD1 Western blot analysis confirmed that the induced cells had remained slender, since much lower levels of the stumpy-specific marker could be detected compared to the arrested uninduced population (Figure 3.2g). The proportion of the induced population in the 1K1N/2K1N configuration remained constant throughout the infection time course, whereas the uninduced cells accumulated in 1K1N as they differentiated in response to SIF (Figure 3.2h).

Over decades of laboratory adaptation, monomorphic cell lines have lost the capacity to respond to SIF and form transmission competent stumpy forms. The *in vivo* data presented here demonstrates that, upon *HYP2* or *DYRK* RNAi induction, cells are unable to fully transduce the SIF signal and have thus become inducibly monomorphic. This manifests as the inability to restrict parasitaemia in response to SIF, the retention of slender morphology and the reduced expression of PAD1 at high parasitaemia. These cell lines were therefore chosen for further use in the study.

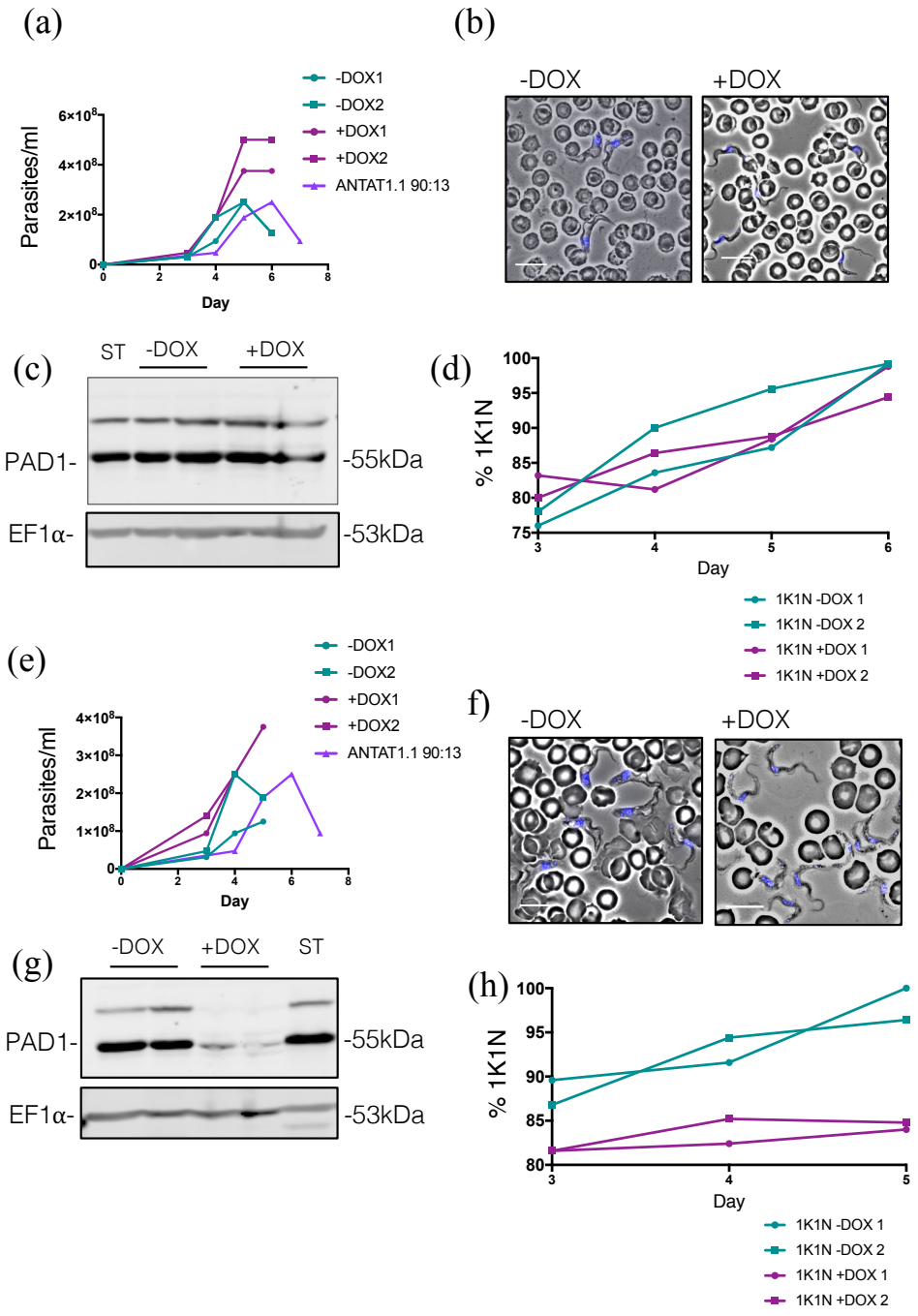


Figure 3.2 Induction of *HYP2* and *DYRK* RNAi reduces differentiation capacity *in vivo*

(a) *In vivo* growth of *HYP2* RNAi cell lines. Sucrose water containing 1µg/ml doxycycline was supplied to two mice from day -1 to induce RNAi. The remaining three mice were supplied with sucrose water alone. The parasitaemia was estimated from a blood smear taken at 24 hour time points from day 3 onwards. The mice were sacrificed before the parasitaemia became lethal and the trypanosomes were purified from the blood on day 6. Induced cells

grew to a higher parasitaemia than the uninduced cells. (b) On day 6 of infection, uninduced cells displayed the typical morphological characteristics of stumpy cells. In contrast, morphologically slender, intermediate and stumpy forms were found in the induced population. Bar, 10µm. (c) Western blot detection of PAD1. Protein was prepared from $4x10^6$ cells which were purified from the blood on day 6. The *T. brucei* AnTat1.1 90:13 parental cells which had arrested as stumpy forms were used as a stumpy form control (ST). Induced cells expressed PAD1 to a slightly lesser extent than the arrested uninduced cells. (d) Cell cycle analysis. Blood smears were fixed in methanol from day 3 onwards for at least 24 hours, then stained with DAPI to visualise the kinetoplast and nucleus. Uninduced cells accumulated in G1/G0, seen as an accumulation of 1K1N cells. Induced cells still accumulated in G1/G0, though at a slower rate than uninduced cells. 250 cells were counted for each replicate. (e) *In vivo* growth of *DYRK* RNAi cells. Infections were performed and monitored as described for the *HYP2* RNAi infection. The mice were sacrificed before the parasitaemia became lethal and the trypanosomes purified from the blood on day 5. The induced cells did not arrest in response to SIF and continued to proliferate. (f) Morphology of methanol-fixed cells from day 5 of infection. RNAi knockdown of *DYRK* caused retention of slender morphology, whilst cells with an intact stumpy formation pathway became morphologically stumpy. Bar, 10µm. (g) Western blot detection of PAD1. Protein was prepared from $4x10^6$ cells isolated on day 5 of infection. Induction of *DYRK* RNAi reduced the expression of PAD1 compared to uninduced cells. (h) DAPI cell cycle analysis. Uninduced cells arrested in G1/G0 as they differentiated, whilst the proportion of induced cells in a 1K1N configuration remained relatively constant. 250 cells were counted per infection.

3.3 VSG AnTat1.1 was the actively expressed VSG in *HYP2* and *DYRK* RNAi cells

Before beginning to design and optimise an assay to interrogate VSG switching, it was important to determine which VSG(s) the *HYP2* and *DYRK* RNAi lines were expressing, and whether the populations were homogenous or heterogeneous with regard to their VSG expression. Identification of the actively expressed VSG would facilitate the generation of both ES targeting and VSG detection strategies. All mature VSG mRNAs are flanked by conserved 5' and 3'

sequences termed the spliced leader (SL) and 14-mer, respectively (Figure 3.3a) (Marcello and Barry, 2007). This allows for the specific amplification of single-stranded VSG cDNA from total RNA using a primer specific to the 14-mer, which itself can then be amplified by conventional PCR using primers targeting the SL and 14-mer.

Total RNA was isolated from $5x10^7$ *HYP2* or *DYRK* RNAi cells which had been purified from the blood in Section 3.2.2. To control for the efficiency of the reverse transcriptase step, an oligo$(dT)_{20}$ primer was used for single-stranded cDNA production instead of the VSG specific 14-mer. This permitted the use of primers amplifying a 500bp fragment of actin in the PCR step. Double-stranded VSG cDNA was generated as described in Section 2.11, and 10µl of product run on an ethidium bromide stained agarose gel. Figure 3.3b shows the results for the *DYRK* RNAi line. The RT-PCR products for the SL and 14-mer primer pair run as a larger ~1600bp fragment, with a faint smaller band running directly below it. Fragments of these sizes are within the 1500-2000bp range that would be expected for an intact VSG, and the presence of two bands most likely represents two different VSGs. The band did not change in size upon the induction of RNAi and nor was it different to the parental AnTat1.1 90:13 population. The band sizes were consistent with those visualised for the *HYP2* RNAi line also (data not shown). The positive control produced the expected 500bp band size. Minus reverse transcriptase samples (-RT) did not produce an RT-PCR product and therefore confirmed that the starting RNA samples were not contaminated with DNA.

The RT-PCR products were purified from the PCR reaction mix before the addition of A-overhangs. These allowed the products to be cloned into the T-overhangs of the pGEM T Easy Vector (Promega). 21 clones were obtained, and their DNA digested with EcoRI (Figure 3.3c). Digest products were run on an agarose gel and stained with ethidium bromide. The digest products of 18 of the 21 clones consisted of a 3kB fragment (digested pGEM T Easy Vector) and two smaller fragments of ~1000bp and 600bp, examples of which are seen in lanes 4 and 6 of Figure 3.3c. The coding sequences for VSG AnTat1.1 and VSG AnTat1.8 have already been cloned, and it was observed that the VSG AnTat1.1 coding sequence

contained an internal EcoRI restriction site (Pays et al., 1980). It therefore seemed likely that the identity of this VSG was VSG AnTat1.1. EcoRI digest of the plasmid DNA acquired from the three remaining clones also produced a 3kB pGEM T Easy fragment, however the cloned fragments were ~1500bp (Lane 2 of Figure 3.3c), 1750bp and 2000bp in length, respectively. The DNA from these three clones, and 5 of the former clones, was sent for sequencing to identify the expressed VSGs. BLAST analysis confirmed that the five sequences obtained for the RT-PCR products which gave the double-band profile upon digestion with EcoRI, and were expressed by the majority of the populations, matched that of VSG AnTat1.1. The sequences obtained for the three clones which did not have internal EcoRI digest sites did not match that of VSG AnTat1.1 nor VSG AnTat1.8, however the SL and 14-mer were identified, therefore confirming that the cloned fragments were really VSG RT-PCR fragments. Immunofluorescence analysis of *HYP2* and *DYRK* RNAi populations stained with an α-VSG AnTat1.1 antibody confirmed the findings of the sequencing and diagnostic digests (Figure 3.3d).

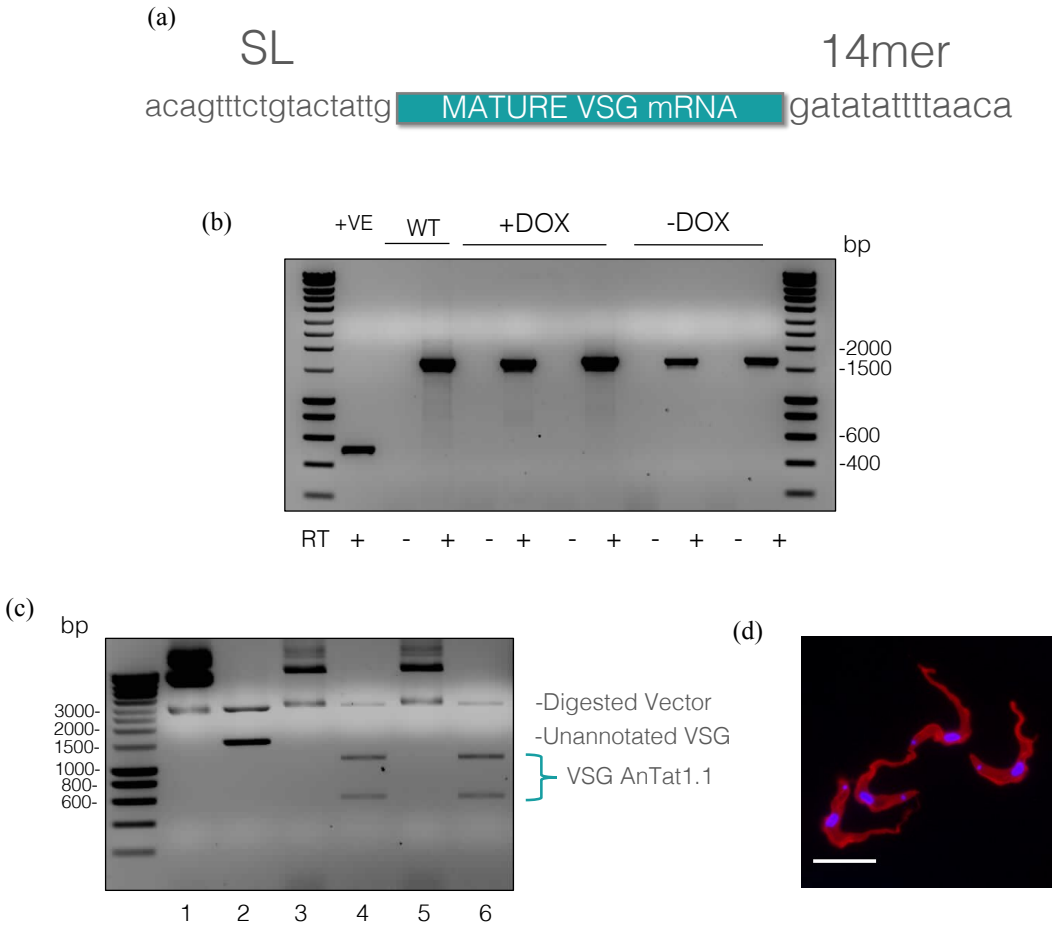


Figure 3.3 Sequencing of RT-PCR products and diagnostic digests identify VSG AnTat1.1 as the primarily expressed VSG in the *HYP2* and *DYRK* RNAi populations

(a) All mature VSG mRNAs are flanked at their 5' and 3' ends by the conserved spliced leader (SL) and 14-mer sequences, respectively. **(b)** RT-PCR amplification of VSGs expressed by the *in vivo DYRK* RNAi populations used in Section 3.2.2. RT-PCR with a SL/14-mer primer pair produced a ~1600bp band with a faint smaller band directly below it. There was no difference in the bands products by the induced, uninduced and parental AnTat1.1 90:13 (WT) populations. RNA which did not have reverse transcriptase (RT) added prior to cDNA amplification yielded no product, therefore confirming that the RNA samples were not contaminated with DNA. Actin was used as the positive control (+VE) for the PCR step. **(c)** RT-PCR products were cloned into the pGEM T Easy vector and the plasmid DNA from resultant clones digested with EcoRI. In addition to the 3kB band representing digested vector, two smaller bands of ~1000bp and 600bp were visualised for 18/21 clones, as seen in lanes 4 and 6. Therefore the cloned VSG RT-PCR products had an internal EcoRI restriction site. Lane 2 contained a different VSG product, which did not have

an internal EcoRI site. The clones whose digest products were ran in lanes 1, 3 and 5 contained only the empty pGEM T Easy vector. **(d)** Induced HYP2 RNAi cells isolated from the final day of an *in vivo* infection stained positive with an α-VSG AnTat1.1 antibody. Bar, 10μm.

3.4 Integration of a GFP reporter within the active VSG expression site

Sequencing of RT-PCR products, antibody staining and diagnostic digests confirmed that the *HYP2* and *DYRK* RNAi populations were largely expressing VSG AnTat1.1, and that induction of RNAi did not induce a widespread antigenic switch. To detect and quantify transcriptional switches away from the AnTat1.1 ES, we integrated a quantifiable reporter at a promoter proximal position of the AnTat1.1 ES. This created a scenario where the α-VSG AnTat1.1 antibody could firstly be used to detect whether a VSG switch had occurred, and secondly a quantifiable reporter could then inform whether this switch had occurred as a result of an *in situ* or recombinogenic switch downstream of the reporter integration site close to the promoter.

The pLF12-eGFP reporter construct (Zimmermann et al., 2017) consists of an upstream promoter region (UP), *eGFP* flanked by the aldolase 5' and 3' UTRs, *PUR* flanked by 5' actin and 3' aldolase UTRs and a downstream promoter region (DP) (Hertz-Fowler et al., 2008). The 5' homologous region, UP, spans the 1590bp region from the 50 bp repeats upstream of the BES promoter to 240bp downstream of the BES promoter and the 3' homologous region, DP, consists of the 1150bp region starting from 287bp downstream of the BES promoter to *ESAG7*. The plasmid can therefore integrate randomly into the promoter proximal region of any BES, however integration of the construct into only the active ES can be forced with high concentrations of the selective drug (Hertz-Fowler et al., 2008). Since the pALC14 RNAi construct itself encodes puromycin resistance, *PUR* was removed from pLF12-eGFP by restriction digest, and replaced by *BLA*, which encodes blasticidin resistance (Figure 3.4a). The plasmid was linearized with SacI and KpnI, and 10μg transfected into the VSG AnTat1.1 expressing *HYP2* and *DYRK* RNAi lines.

Selection with 10µg blasticidin yielded the GFP^{ESpro}AnTat1.1^{ES} *HYP2*/*DYRK* RNAi cell lines (Figure 3.4b). When the VSG AnTat1.1 expression site was active the cells emitted a homogenous cytoplasmic GFP signal which was not lost following fixation (Figure 3.4c). Populations of clones were ~99% VSG AnTat1.1 and GFP positive, as measured by FACS (Figure 3.4d).

In addition to the *HYP2* and *DYRK* RNAi lines, a GFP ORF was also transfected into the VSG AnTat1.1 ES of a pleomorphic *NEK17* RNAi cell line. NEK17 (herein referred to as just 'NEK') is encoded by three tandem genes (Tb927.10.5930/40/50) which are indistinguishable by RNAi. *NEK* was identified by Mony et al. (2014) in their genome-wide RNAi screen as a component of trypanosome quorum-sensing, and inducible knock-down was shown to ablate differentiation competence in pleomorphic cells *in vivo*, therefore making the molecule a potential stumpy inducer (Mony et al., 2014). McDonald et al. (2018) more recently demonstrated that ectopic overexpression of NEK was not able to restore differentiation competence of AnTat1.1 90:13 *RBP7* (another stumpy inducer) or *DYRK* null mutants *in vivo*, thus demonstrating that NEK driven stumpy formation is both RBP7 and DYRK dependent (McDonald et al., 2018).

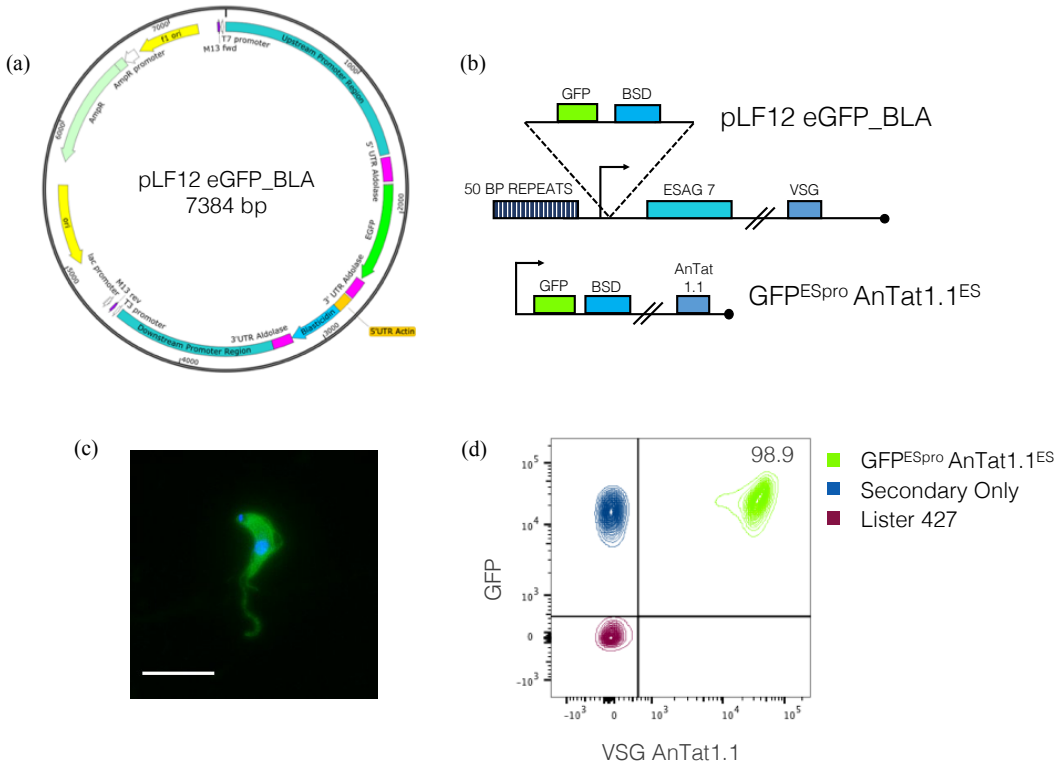


Figure 3.4 GFP^{ESpro}AnTat1.1^{ES} H*YP2*/*NEK*/*DYRK* RNAi cell lines express GFP from their VSG AnTat1.1 ES (a) pLF12 eGFP_*BLA* encodes eGFP and blasticidin resistance **(b)** pLF12 eGFP_*BLA* targets the ES promoter region. The construct integrated into the actively expressed VSG AnTat1.1 ES of the RNAi cell lines, generating the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi reporter cell lines **(c)** When the VSG AnTat1.1 ES was active, the cells emitted a strong cytoplasmic GFP signal. Bar, 10µm **(d)** FACS analysis of 10,000 paraformaldehyde/glutaraldehyde fixed cells demonstrated that the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi reporter cell lines were 98.9% VSG AnTat1.1 and GFP positive. VSG AnTat1.1 positivity was measured by α-VSG AnTat1.1 antibody binding and GFP positivity by fluorescence emission in the FITC channel alone.

3.5 The GFP^{ESpro} AnTat1.1^{ES} *HYP2*/*NEK*/*DYRK* RNAi cell lines remained inducible and lost differentiation competence upon induction

To validate that the insertion of an eGFP ORF into the VSG AnTat1.1 ES had not disrupted the functionality or inducibility of the RNAi machinery, duplicate

mouse infections were established (± induction) with the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi reporter lines as in Section 3.2.2. Induced RNAi cell lines did not restrict their parasitaemia in response to accumulating levels of SIF, and furthermore retained their slender morphology (Figure 3.5a). Uninduced cells, in contrast, committed to differentiation and transformed morphologically to stumpy cells from day 4 of infection onwards, with similar kinetics to the parental *T. brucei* AnTat1.1 90:13 cell line. Northern blots confirmed efficient and inducible RNAi mediated knockdown of the *HYP2* and *DYRK* transcripts (Figure 3.5a, inset). *NEK* transcripts were knocked down to a much lesser degree, however since hyperparasitaemia and the retention of slender morphology were nevertheless observed in the induced lines, this suggested that only small levels of gene knockdown were required to significantly reduce differentiation competence in these lines. Cell-cycle analysis of 250 DAPI stained-cells from each day of infection revealed a reduced accumulation of 1K1N cells in the induced RNAi lines compared to the uninduced, indicating a loss or delay in the G1/G0 cell cycle arrest associated with the differentiation to stumpy forms (Figure 3.5b). Consistent with the reduced G1/G0 arrest and maintenance of slender morphology, induced cells isolated on the final day of infection expressed the stumpy specific marker PAD1 to a much lesser extent compared to the uninduced cells (*HYP2* RNAi) or not at all (*NEK* and *DYRK* RNAi) (Figure 3.5c). Together, this data confirmed that the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines lost differentiation competence upon induction, and were therefore inducibly monomorphic, as demonstrated by the retention of slender morphology, reduced G1/G0 arrest at high parasitaemia and diminished expression of the stumpy form marker, PAD1.

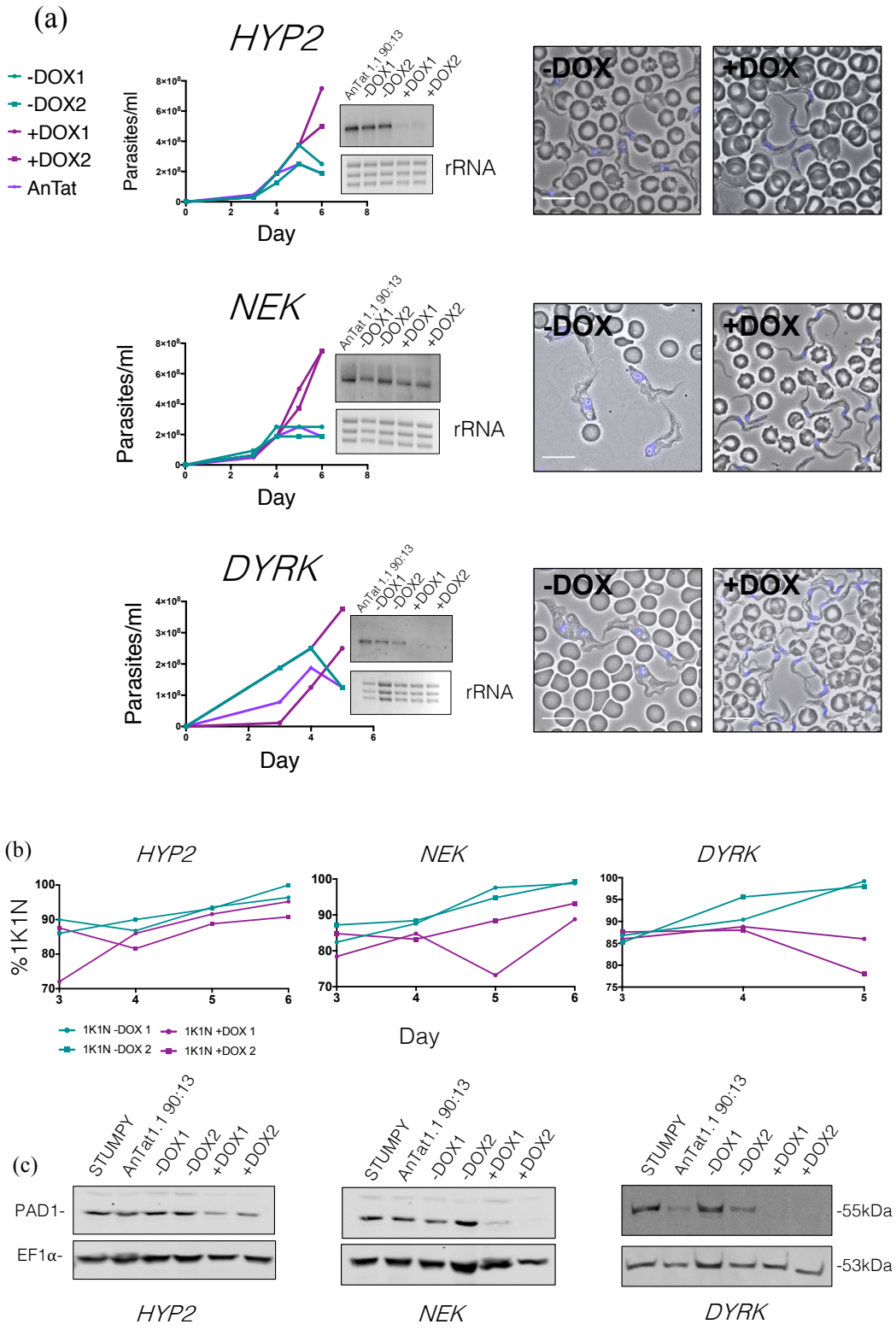


Figure 3.5 The GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines are inducibly monomorphic (a) *In vivo* growth and final day cell morphology of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* (top), *NEK* (middle) and *DYRK* (bottom) RNAi cell lines. The mice were sacrificed

before the parasitaemia became lethal and the trypanosomes were purified from the blood on day 5 (*DYRK* RNAi) or day 6 (*HYP2* and *NEK* RNAi). Induction of RNAi generated populations of cells which grew to high parasitaemia and were slender in morphology. Uninduced populations of cells restricted their parasitaemia from day 4 onwards and were enriched for morphologically stumpy forms. Northern blot detection (inset) of *HYP2*, *NEK* and *DYRK* transcripts revealed varying degrees of gene knockdown upon induction with doxycycline. Bar, 10μm **(b)** Cell cycle analysis of $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2* (left), *NEK* (middle) and *DYRK* (right) RNAi cell lines. Uninduced cells accumulated in G1/G0 as they differentiated to stumpy forms, seen as an accumulation of 1K1N cells. The percentage of replicative cells remained constant in the induced *NEK* RNAi population and increased upon induction of *DYRK* RNAi. Induced *HYP2* RNAi cells still accumulated in G1/G0, though at a slower rate than uninduced cells. 250 cells were counted for each replicate. **(c)** Western blot detection of PAD1 protein. Protein was prepared from $4x10^6$ cells which were purified from the blood on the final day of infection. The *T. brucei* AnTat1.1 90:13 parental cells which had arrested as stumpy forms were used as a control (ST). Consistent with their differentiation to stumpy form cells, the uninduced populations expressed high levels of PAD1. In contrast, the induced *HYP2*, *NEK* and *DYRK* RNAi cells showed diminished PAD1 expression.

3.6 Optimisation of an *in vitro* flow cytometry-based assay to measure VSG switch rate

The $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines are inducibly monomorphic and contain a promoter-proximal GFP ORF in their active expression site, therefore permitting the study of antigen switch frequency and mechanism in the same cell lines where differentiation capacity is switched 'on' or 'off'. VSG AnTat1.1 positivity can be measured using an α-VSG AnTat1.1 antibody and the promoter proximal GFP ORF allows for the detection of ES transcriptional switches. Many factors had to be considered when designing a method to interrogate VSG switching: how should GFP and VSG AnTat1.1 positivity be detected; should the assay be performed *in vitro* or *in vivo*; what scale should the assay be performed on; how should the populations of cells be maintained in logarithmic growth; how do we avoid the propagation of pre-

existing switched clones in the parental population; and how should data be interpreted and VSG switch rates calculated?

3.6.1 Designing an *in vitro* VSG switch assay workflow

Early studies of trypanosome antigenic variation typically monitored variants arising from single relapses in rodents (Miller and Turner, 1981). Variations on this technique included immunising mice against a specific (and predominantly expressed) VSG and then detecting an emerging parasitaemia after infection (Hartley and McCulloch, 2008; McCulloch and Barry, 1999; McCulloch et al., 1997; Proudfoot and McCulloch, 2005; Robinson et al., 2002) or depleting a population of trypanosomes expressing a specific VSG *in vitro* by antibody-mediated lysis in serum and then inoculating the cells into mice and detecting infection (Lamont et al., 1986; Turner, 1997). Whilst single relapse experiments account for the potential differences in growth rates between variable antigen type (VAT) variants and make the exact parentage of a VSG switched cell clear, *in vivo* VSG switch assays often require large numbers of rodents to be screened and prior immunisation of animals (Aitcheson et al., 2005). Furthermore, antibodies could clear all VSGs expressing similar epitopes to the variant that was immunised against, potentially leading to an underestimation of VSG switch rate. During our *in vivo* differentiation experiments the uninduced RNAi cell lines committed to differentiation by day 4, as demonstrated by the restriction of parasitaemia and the accumulation of 1K1N cells (Figure 3.5). To perform a valid comparison of VSG switch rates, both uninduced and induced populations therefore need to be maintained in logarithmic growth (this is because arrested stumpy cells do not contribute new antigenic variants to the population (Amiguet-Vercher et al., 2004)). Controlling the growth of trypanosome populations to defined densities is much more achievable in a culture model than in animals and therefore, with this and the prior arguments in mind, it was decided to design an *in vitro* VSG switch assay.

Many *in vitro*-based VSG switch assays have been developed to circumvent the limitations and constraints presented by *in vivo* methods (see Section 1.2.7). Typically, these experiments consist of a non-selective growth period, in which variants arise, followed by a step which depletes unswitched cells from the

population, for example MACS (Boothroyd et al., 2009; Schulz et al., 2016), RNAi against the original VSG (Aitcheson et al., 2005) or Ganciclovir treatment to remove cells expressing the HSV-TK gene from the original ES (Devlin et al., 2016; Kim and Cross, 2010). This depletion/enrichment step is performed to increase the depth of analysis since the laboratory-adapted monomorphic trypanosomes which have been used for these studies switch their expressed VSG at rates between 10^{-7} and 10^{-5} switches/cell/generation. Following the depletion of cells that have not switched VSG expression, the remaining variants are typically cloned in 96 well plates by limiting dilution and then characterised by methods such as IFA, PCR and Southern blotting.

Pilot experiments with the inducibly monomorphic RNAi lines suggested that 1 in every 100 trypanosomes switched their expressed VSG in the absence of selection. Since VSG switching appeared to be occurring at a high frequency, and to avoid unintentional bias, the decision was taken not to deplete the RNAi lines of unswitched cells following the initial period of *in vitro* growth. The use of IFA, PCR or Southern Blots to characterise VATs is time-consuming and labour-intensive, limiting the number of variants that can be realistically studied. Flow cytometry-based analysis of switch events is advantageous since it firstly allows for the rapid processing of millions of cells and, secondly, can detect multiple fluorescent markers simultaneously. With the correctly placed markers, flow-cytometry can not only identify VSG switches but also report on the method of VSG switching. It was thus decided that a flow cytometry-based strategy for the detection of VSG AnTat1.1 and GFP positivity should be used.

With regard to deciding the scale of the assay, we know that the collection of 10,000 events during flow cytometry gives a confidence interval of 99% (Hedley and Keeney, 2013). Therefore, to demonstrate that any *in vitro* generated VSG switch events recorded were ‘real’, a VSG switch assay sample would have to contain at least 10,000 switched cells. Following pilot *in vitro* experiments, ~1 in every 100 trypanosomes were consistently found to be VSG AnTat1.1 negative. Therefore, to measure 10,000 VSG AnTat1.1 negative events, 1,000,000 total events for every switch assay replicate would have to be recorded by the flow cytometer. To

compensate for loss of sample during fixation and antibody staining, $2x10^7$ cells were taken from each FACS switch assay for fixation. *In vitro*, pleomorphic trypanosomes should be maintained below cell densities of $1x10^6$ cells/ml to prevent initiation of slender to stumpy differentiation. Therefore, the VSG switch assay would have to finish with at least 20ml of trypanosomes at a cell density of $1x10^6$ cells/ml to produce enough material for analysis.

To guarantee that recorded VSG switches were independently generated during the switch assay, and not a result of the amplification of pre-existing variants within the parental population, the assay would also have to begin with either a homogenous population of cells expressing the dominant VSG (VSG AnTat1.1), or a single cell expressing a known VSG. Since earlier FACS analysis had demonstrated that the GFP^{ESpro}AnTat1.1^{ES} *HYP2*/*NEK*/*DYRK* populations were not homogenous with regard to GFP and VSG AnTat1.1 expression (Figure 3.4d), and to avoid an additional FACS-sorting step prior to the switch assay commencing, a strategy where a single VSG AnTat1.1/GFP positive cell was 'scaled-up' in culture was adopted. Briefly, a logarithmic culture of trypanosomes was diluted and distributed over a 96-well plate such that less than one cell/100µl well was seeded. Cells were cultured in non-selective medium to allow for *in situ* ES switching. Following 5 days incubation at 37^{o}C, clones were counted with a haemocytometer and three selected to be transferred into 2ml of medium. This procedure was repeated with increasing volumes of HMI-9 until at least $2x10^7$ cells could be collected for fixation. This scaling up procedure assured that firstly, the populations of cells did not grow $>1x10^6$ cells/ml and secondly, that the expressed VSG population was not restricted by bottlenecking. Doxycycline was added to the parental culture just prior to dilution, and at every scale up step thenceforth. The basic workflow is shown in Figure 3.6. On average, the uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2*/*NEK*/*DYRK* RNAi VSG switch assays lasted 10 days (range=3, n=18), encompassing 26.93 population doublings (± 0.56, n=18). The induced assays also lasted, on average, 10 days (range=3, n=17), encompassing 26.69 population doublings (± 0.81, n=17). Since the calculated VSG switch rates were normalised to the number of population doublings (see Section 3.6.3) it was not considered problematic that the VSG switch assays did

not always run for exactly the same duration. What was imperative was that the uninduced and induced populations were maintained in logarithmic growth.

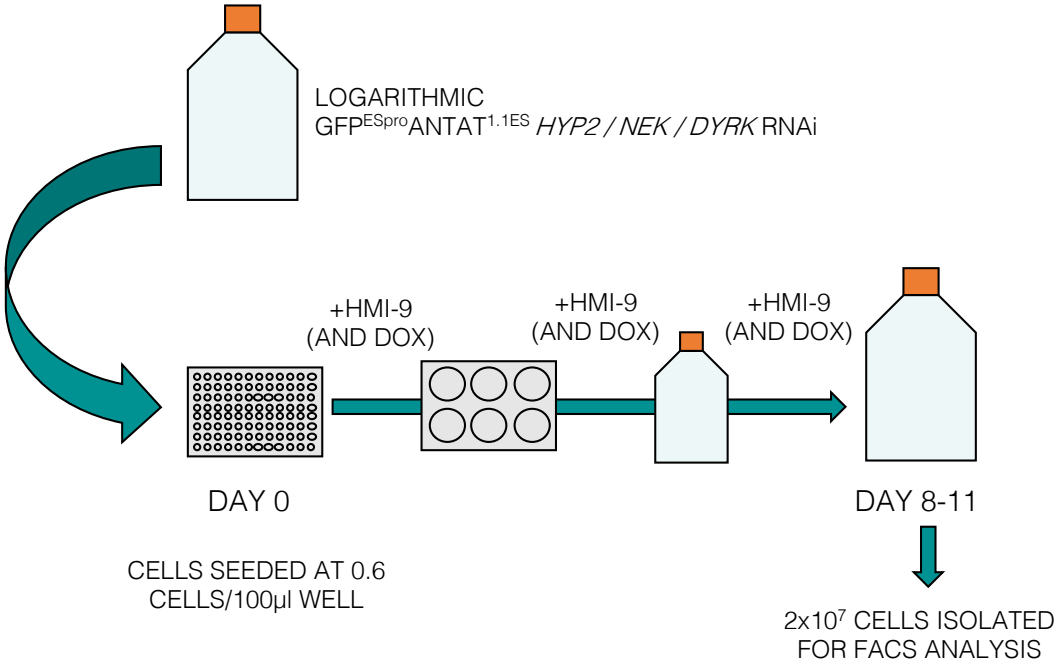


FIGURE 3.6 *In vitro* FACS-based VSG switch assay workflow 10ml of a logarithmic uninduced population of GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* or *DYR*K cells was diluted to 6 cells/ml and seeded in 100µl volumes in a 96-well plate. Doxycycline was added to the +DOX populations following dilution, and prior to distribution in the 96 well plate. From the remainder of the starting logarithmic population, 2x10^7 cells were fixed and stored for use as a FACS positive control/gating aid. After 5 days of incubation at 37^{o}C, the cell density of the arising clones was counted on a haemocytometer. Typically, seeding at 0.6 cells/well yields 3-20 clones per plate. Three clones that were at similar cell densities, and appeared healthy when visualised with a microscope, were selected for progression through the assay. Over the course of a further 3-6 days, cell counts were taken daily and the cultures progressively scaled up by the addition of HMI-9 to prevent the populations from overgrowing. When fresh HMI-9 was added to the induced cultures, doxycycline was 'topped up' too. When at least 2x10^7 cells were acquired, the growth phase of the assay was ended and cells washed and fixed before FACS analysis. Where possible, the volume of the final culture was kept as low as possible so that as high a proportion of the population as possible was removed for FACS analysis.

3.6.2 Interpreting the FACS data

Prior to fixation, cells were initially washed and incubated with the fixable live/dead stain, Zombie, (Biolegend) which reacts with the primary amine groups on proteins. When live cells are treated, the plasma membrane excludes the entry of the dye meaning that only the surface proteins will react. However, in dying cells the loss of the membrane integrity means that the dye can enter the cell and therefore bind both extra- and intra-cellular proteins. These dead cells will hence fluoresce more strongly than live or unstained cells. The inclusion of this dye meant that GFP negativity brought about by cell death could be excluded from the analyses. Treatment with the Zombie dye, rather than conventional propidium iodide treatment (PI), was selected since the staining occurs before, not after, fixation. Therefore, the cellular integrity of the population prior to the potentially damaging fixation process could be investigated. Cells were stained with the α-VSG AnTat1.1 antibody as described in Section 2.16.2. Immediately before processing samples on the flow cytometer, a defined number of CountBright absolute counting beads (Thermofisher Scientific) were added to each of the switch assay samples. As the stained cells were processed by the flow cytometer, so too were the counting beads. Since the exact number of beads added to the FACS sample was known, the absolute number of cells in a sample could be determined by relating the number of cells acquired by the flow cytometer to the number of processed bead events, and then multiplying by the culture dilution factor (Figure 3.7) (Schulz et al., 2016). The quantification of VSG switchers and the subsequent calculation of VSG switch rate is described in detail in Section 3.6.3.

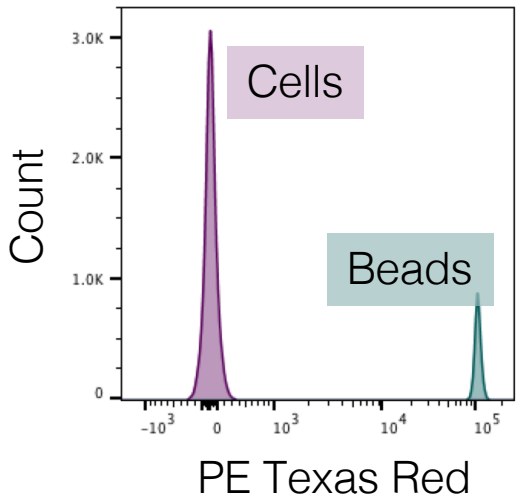


Figure 3.7 Counting beads gating strategy Countbright absolute counting beads fluoresce brightly over a wide range of excitation (UV to 635nm) and emission (385-800nm) wavelengths. The beads were separated from the population of cells by gating in the PE Texas Red (596nm excitation and 615nm emission) channel. At least 1000 beads were collected per sample for statistical significance.

One million events were acquired for each triplicate sample by the BD LSRII flow cytometer (BD Biosciences). Before determining VSG AnTat1.1 and GFP positivity, the flow cytometry data was filtered using the forward (FSC) and side scatter (SSC) profiles, such that only intact cells that were processed singularly were selected for analysis (Figure 3.8a, panels 1 and 2). Next, single cells were observed for Zombie dye positivity (Figure 3.8a, panel 3) with the intention of removing dead cells from the population. However, such was a) the integrity of cells following the assay, and b) the high stringency of gating for intact and single cells that this stage consistently failed to exclude any more than a couple of cells from the population in trial experiments. GFP negative events could therefore not be attributed to the loss of GFP expression upon cellular death. Since dead cells were so rarely found in the populations, the use of the expensive Zombie dye was excluded from the final analyses.

The next gate to be applied to the population was that determining VSG AnTat1.1 positivity (Figure 3.8b). To aid gating, a VSG 221 expressing Lister 427

clone and the (largely) VSG AnTat1.1 expressing logarithmic population from which the switch assay was started were used as negative and positive controls, respectively, in addition to a VSG AnTat1.1 expressing secondary antibody-only control. Following a genetic VSG switch, trypanosomes require ~4.5 days to completely replace their surface coat (Pinger et al., 2017). Therefore, whilst the VSG AnTat1.1 signal from cells that had switched early in the assay would match that of the negative and secondary antibody-only controls, it was likely that some of the VSG switched cells would only lose a small proportion of their VSG AnTat1.1 positivity. Hence, positioning a gate directly to the right of the negative and secondary only controls would wrongly classify cells that had undergone a transcriptional or genetic VSG switch, but replaced little of their VSG surface coat, as VSG AnTat1.1 positive. Consequently, the gate defining VSG AnTat1.1 positivity was placed directly to the left of the positive control. Thus, cells that had switched expression away from VSG AnTat1.1, but had only lost a very small proportion of their VSG AnTat1.1 surface protein, would still be classed as switchers.

Cells that were defined as VSG AnTat1.1 negative (circled in Figure 3.8b) were next analysed for GFP positivity to determine the mechanism of VSG switch. ES transcription has been demonstrated as initiating on multiple silent ESs and progressing at least as far as, and including, ESAG 6 before the arrest of RNA elongation in all but the active ES occurs (Vanhamme et al., 2000). On an unmodified AnTat1.3A ES, ESAG 6 is positioned ~4kbp downstream of the ES promoter. That transcription progresses beyond the GFP reporter, which is positioned 1250bp downstream of the AnTat1.3A ES promoter in the GFP reporter RNAi cell lines, could have implications for the interpretation of VSG switch mechanisms. Specifically, if a cell were to switch transcription away from the AnTat1.3A ES and activate another ES, transcription occurring on the now silenced AnTat1.3A ES could still produce a positive GFP signal. Selection against transcription on silent sites by the addition of high levels of selective drug (Hertz-Fowler et al., 2008) is not preferable since this would either select against *in situ* switching, or rapidly kill any parasite that had. Transcription on silent ESs however, seems to be at a significantly lower level that that of the active ES, suggesting that GFP mRNA levels would decrease upon an *in situ* switch away from the AnTat1.3A ES. Vanhamme et al.

(2000) investigated transcript heterogeneity from 4 regions of active and inactive ESs by RT-PCR: (i) immediately adjacent to the ES promoter, (ii) upstream of ESAG 7, (iii) a conserved region within ESAG 7 and ESAG 6, and (iv) the VSG itself. In a VSG AnTat1.1 expressing clone, the transcripts were predominantly from the AnTat1.3A ES. However, 4% and 3% transcript heterogeneity was observed from the region immediately downstream of the promoter and the ESAG7/6 segment, respectively, indicating that whilst transcription on silent sites was occurring, it was greatly attenuated compared to the active ES (Vanhamme et al., 2000). Using the same pLF12 eGFP plasmid that has been adopted for this study, Zimmerman et al. (2017) furthermore demonstrated mRNA levels of the promoter proximal eGFP decreased by 80% in the 24 hours following induction of ES attenuation. Whilst this data suggests that transcription of GFP mRNA would be rapidly attenuated, this FACS-based assay measures fluorescence from the GFP protein. Although the protein itself has been estimated to have a half-life of ~80 hours in *Leishmania infantum* promastigotes (Kamau et al., 2001) and ~26 hours in mammalian cells (Corish and Tyler-Smith, 1999), Aresta-Branco et al. (2016) demonstrated by FACS that trypanosomes expressing GFP immediately downstream of their active VSG ES promoter became GFP negative 120 hours after transcriptional shut down of their active ES (earlier timepoints were not measured) (Aresta-Branco et al., 2016). Together this data suggests that GFP mRNA is rapidly turned-over but a GFP signal could still be detected up to 5 days post-transcriptional switch, albeit at a lower level. Therefore, the gate defining GFP positivity was again placed directly to the left of the ‘Day 0’ GFP positive control (Figure 3.8c). A cell line which had not been transfected with GFP served as a negative control.

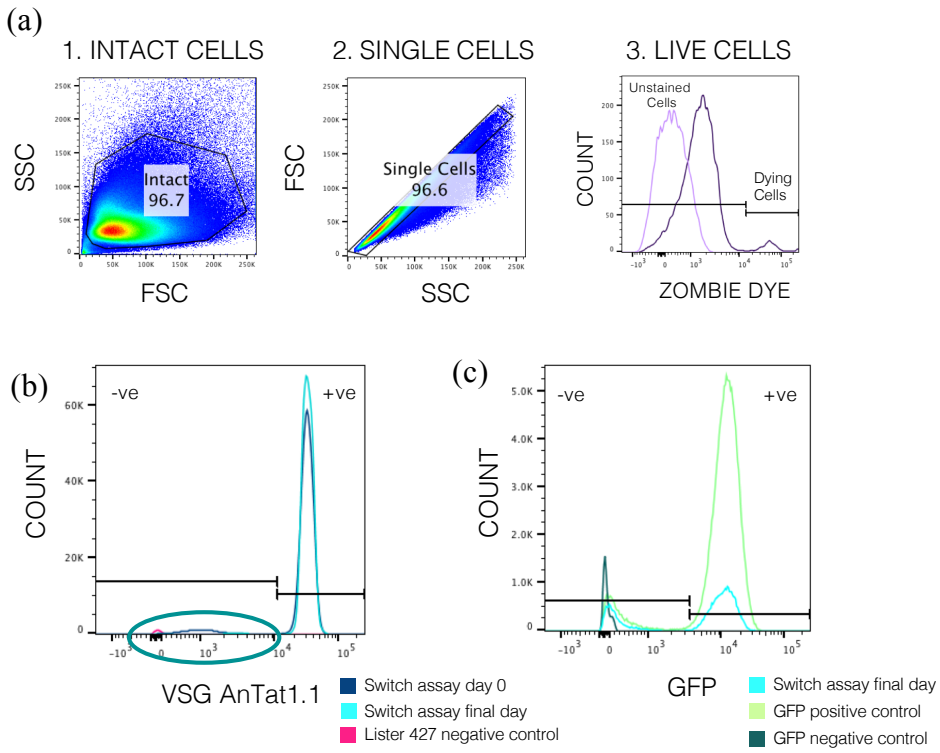


Figure 3.8 VSG AnTat1.1 and GFP gating strategies (a) Debris and cell doublets were removed from the population using the FSC and SSC profiles. In trial experiments, cells were stained with the fixable live/dead Zombie stain prior to fixation and antibody staining. The first two gating stages were sufficient to remove dead cells and hence, use of the dye was excluded from the final experiments. (b) Cells were first analysed for VSG AnTat1.1 positivity based on α-VSG AnTat1.1 staining. The gate was placed directly to the left of the VSG AnTat1.1 positive peak to account for cells taking up to 4.5 days to fully replace their surface coat following a genetic VSG switch. A VSG221 expressing Lister 427 clone was used as a negative control and a VSG AnTat1.1 expressing clone was used for the secondary antibody-only control (not shown on the plot). The logarithmic population from which the VSG switch assay was started was used as the VSG AnTat1.1 positive control. (c) VSG AnTat1.1 negative cells were analysed for GFP positivity to determine the mechanism of VSG switch. The GFP positive gate was placed directly to the left of the GFP positive peak to account for the half-life of the protein. A cell line which had not been transfected with GFP was used as a negative control. Since

trypanosomes autofluoresce in the FITC channel, the mean GFP fluorescence of the negative population is always higher than zero. The population used for the VSG AnTat1.1 positive control was also used as the GFP positive control.

3.6.3 Calculating the VSG switch rate

To extrapolate the VSG switches/cell/generation from the gated FACS plots, a method based on the calculations described by Schulz et al. (2016) was adopted (Table 3.1). For the following calculations, 'sample' refers to the cells that were removed from the VSG switch assay on the final day, fixed, stained and analysed on the flow cytometer. 'Population' refers to the cells contained within the culture flask on the final day of the switch assay. The number of beads and the number of intact single cells acquired for each sample were entered in columns 2 and 3 of the table, respectively. The number of counting beads that were added to the FACS sample prior to sorting were then entered in column 4. Since,

$$\frac{Number\ of\ beads\ acquired}{Number\ of\ cells\ acquired} = \frac{Total\ number\ beads\ added}{Total\ number\ cells\ in\ sample}$$

The absolute number of cells at the end of the VSG switch assay was calculated as,

$$\frac{Total\ beads\ added\ \times\ Cells\ acquired}{Beads\ acquired} \times Dilution\ Factor$$

This is shown in column 5 of Table 3.1. The dilution factor refers to the proportion of the final VSG switch assay culture that was removed and fixed for FACS analysis. For example, if half of the final culture was removed and fixed for FACS analysis, then the equation on the left was multiplied by 2. The number of cells gated into the VSG AnTat1.1 negative population (Figure 3.8b, circled) was entered into column 6. At this point of analysis, GFP positivity or negativity was not

considered since it was only an indicator of switch mechanism, not frequency. The absolute number of VSG AnTat1.1 negative cells in the population was then calculated in the same way as was done for the absolute number of cells in the population, replacing 'total cells acquired' for 'total switchers acquired'. This was plotted in column 7. By dividing the absolute number of VSG AnTat1.1 negative cells in the population by the absolute number of cells in the population, the proportion of VSG AnTat1.1 negative cells in the population was calculated (column 8). To convert this to VSG switches/cell/generation (column 10), the proportion generated in column 8 was divided by the total number of population doublings that were counted during the VSG switch assay (column 9). As acknowledged in Section 1.2.7, this calculation did not account for the phenotypic or genetic diversity of the 'switched' population, nor potential differences in growth rate between different variants, it simply presented the proportion of the population at the time of sampling that did not express the initiating VSG (Hovel-Miner et al., 2012). This is problematic when calculating absolute switch frequency. For the nature of these experiments however, i.e. a comparison of two populations prior to and after induction of RNAi, a comparison of the relative VSG switching frequencies suffices. It was assumed that an increase in VSG switch frequency would result in a higher proportion of VSG negative cells within the population, and vice versa.

To determine the percentage of VSG switches that were generated by DNA recombination, the number of VSG AnTat1.1 negative cells in the sample that gated as GFP positive (Figure 3.8c, right gate) was divided by the total number of VSG AnTat1.1 negative cells that were acquired in the sample.

Table 3.1 VSG switch rate calculations Based on Schulz et al. (2016)

	1. Sample name	2. Beads acquired in sample	3. Live cells acquired in sample	4. Beads added to sample	5. Total cells in population	6. Switchers acquired in sample	7. Total switchers in population	8. VSG switches/cell	9. Population doublings	10. VSG switches/cell/ generation
					(live cells acquired*beads added/beads acquired)*df	VSG AnTat1.1 -ve	(switchers acquired*beads added/beads acquired)*df	total switched in population/Total cells in population		(VSG switches/cell)/ (population doublings)
HYP2 RNAi -DOX	1	27488	1163273	4.32E+04	2851982.47	59357	145524.85	5.10E-02	26.83	**1.90E-03**
	2	27917	1161711	4.32E+04	2876292.74	31419	77790.64	2.70E-02	26.80	**1.01E-03**
	3	17180	1134215	4.32E+04	7130105.94	38452	241723.86	3.39E-02	27.44	**1.24E-03**

3.7 Validating the flow cytometry approach as a suitable method to measure VSG switch rate

Before investigating whether the induction of monomorphism caused a reduction in VSG switch frequency, I first sought to validate (a) the ability of the VSG switch assay to detect change in VSG switch frequency, (b) the ability of the switch assay to correctly determine VSG switch rate, and (c) the ability of the switch assay to discriminate between *in situ* and DNA recombination-based switches.

3.7.1 The FACS-based VSG switch assay detected induced VSG switches and could determine the mechanism of VSG switching

To test first, the ability of the assay to detect a change in VSG switch rate, and secondly, how accurately it could predict the type of VSG switch mechanism, the VSG^{up} strain was acquired from the Horn Lab (Dundee, UK). The VSG^{up} strain contains an I-*Sce*I restriction site and the puromycin *N*-acetyltransferase (*PAC*) gene adjacent to the 70 bp repeats upstream of VSG 221 (Figure 3.9a, top) (Glover et al., 2013a; Glover et al., 2013b). A tetracycline inducible I-*Sce*I meganuclease ORF is furthermore integrated into a silent RRNA spacer locus (Glover et al., 2007). Upon the induction of I-*Sce*I expression by the addition of doxycycline, the meganuclease recognizes its 18bp cleavage site in the ES and generates a double-stranded break (DSB). This break is repaired through 70 bp repeat recombination, thus translocating a new, antigenically distinct VSG into the ES. Transcription is not altered. The VSG^{up} strain was therefore useful as it tested whether the VSG switch assay detected the increase in VSG switching upon induction and, when GFP was transfected into the VSG 221 ES, it tested whether the assay correctly identified the recombination-mediated VSG switches as VSG 221 negative and GFP positive (Figure 3.9a). The pLF12 eGFP_BLA construct contained an I-*Sce*I cleavage site upstream of the eGFP start codon (Figure 3.9b). Induction of DSBs in the ES promoter proximal region have been shown not to induce VSG switches (Glover et al., 2013a). To avoid cleavage at two sites upon induction, site directed mutagenesis was performed to mutate two bases at either side of the I-*Sce*I cleavage site in the pLF12 eGFP_BLA

plasmid (Figure 3.9c, top). A diagnostic digest with recombinant I-*Sce*I and BamHI confirmed that the meganuclease no longer cut at the site (Figure 3.9c, bottom).

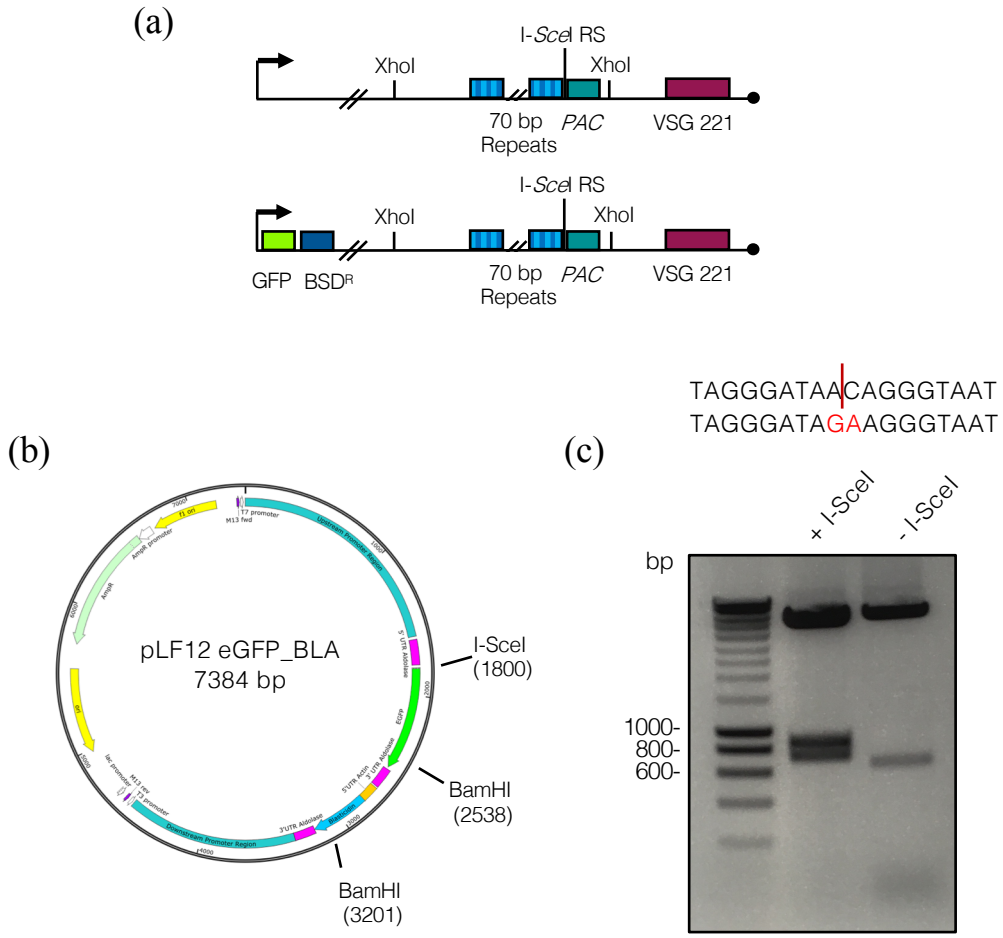


Figure 3.9 Transfection of the GFP reporter construct into the active VSG221 ES of the VSG^{up} cell line (a) Top: The VSG 221 ES of the VSG^{up} strain contained an I-*Sce*I recognition sequence (RS) and the *PAC* gene adjacent to the 70 bp repeats. Tetracycline-inducible I-*Sce*I was encoded from a RRNA spacer locus. **Bottom:** Transfection of the VSG^{up} strain with the pLF12 eGFP-BLA construct allowed us to determine if the assay correctly identified DNA recombination. **(b)** pLF12 eGFP_BLA encoded an I-*Sce*I cleavage site upstream of the eGFP start codon. **(c) Top:** The 5' 18bp sequence recognised by the I-*Sce*I meganuclease before, and after, site directed mutagenesis. The red line denotes where the cut is made. The two mutated bases are shown in red. **Bottom:** Plasmid DNA from the pLF12 eGFP_PUR plasmid (+I-*Sce*I) and the pLF12 eGFP_BLA plasmid after mutagenesis

(-I-*Sce*I) was digested with recombinant I-*Sce*I and BamHI. When the I-*Sce*I cleavage site was present, the digest produced three products, as opposed to two when the site was absent. Following mutagenesis, the digest of the plasmid produced two bands, indicating that the meganuclease no longer recognised the digest site. The expected band sizes for the pLF12 eGFP_PUR plasmid are 5983bp + 738bp + 863bp. The expected band sizes for the mutated pLF12 eGFP_BLA plasmid are 6271bp +663bp.

The VSG^{up} strain was transfected with the mutated pLF12 -I-*Sce*I eGFP_BLA plasmid and three VSG 221/GFP positive clones generated following selection with 10μg/ml blasticidin. To test whether an enrichment for VSG 221 negative/GFP positive cells was detected by the FACS assay, the three clones were induced for three days with 1μg/ml doxycycline (Figure 3.10a). Since it had been demonstrated that only ~5% of VSG^{up} cells survive upon induction of I-*Sce*I expression (Glover et al., 2013a), cultures were initiated at 10^4cells/ml to produce enough material for analysis. The populations of uninduced and induced cells were taken from the same starting logarithmic culture of uninduced cells. On day 3 of the assay, ~$1.5x10^6$ cells from the uninduced and induced cultures were washed and fixed for FACS analysis. Cells were stained with an α-VSG 221 antibody (courtesy of Monica Mugnier, Baltimore, USA) and analysed for VSG and GFP positivity as described in Figure 3.8. 10,000 events were acquired by the flow cytometer.

Following three days of uninduced growth without meganuclease production, the populations of Clones 1, 2 and 3 were largely VSG 221 and GFP positive (77.4 ± 8.5%, 90.6 ± 3.1% and 88.4 ± 6%, respectively) (Figure 3.10b and c). Since the starting cultures were not homogenous for GFP and VSG 221 expression it could not be determined how many of the VSG 221 negative cells arose from independent switch events, as opposed to the propagation of a pre-existing VSG 221 negative cell. Upon induction, there was a clear shift away from VSG 221 expression, such that the populations of Clones 1, 2 and 3 were significantly enriched ($p<0.05$, student's t-test) for VSG 221 negative cells with mean fold changes of 3.1, 7.7 and 5.7, respectively (Figure 3.10c). Upon the induction of I-*Sce*I expression in the parental VSG^{up} line, Glover et al. (2013) reported that the VSG switches/cell increased 5000-fold compared to the value of ~$1x10^{-5}$ switches/cell/generation estimated in the literature

for laboratory-adapted monomorphs, however they never tested the switch rate of the uninduced cell line itself.

Expression of the I-*Sce*I meganuclease, and formation of a DSB adjacent to the 70 bp repeats induces a change in VSG expression via recombination through the 70bp repeats. Therefore, in the induced populations, cells should stain as VSG 221 negative but remain GFP positive. In the three induced clones, the populations of VSG 221 negative cells were 72.2 ± 1.7%, 71.5 ± 2.3% and 69.2 ± 5.6% positive for GFP expression, such that the majority of induced switches had been generated by DNA recombination (Figures 3.10b and d). Whilst significantly more VSG221 negative cells were detected in the induced populations overall, the proportion of VSG 221 negative cells which remained GFP positive was not significantly different ($p>0.05$, student's t-test) between the uninduced and induced populations for clones 1 and 2 (Figure 3.10d). This suggests that the VSG switches detected in the uninduced populations could have been the result of leaky I-*Sce*I expression. As can be seen in Figure 3.10b, the number of VSG 221 negative cells that were also GFP negative increased upon induction too, however the fold-changes (2, 3.9 and 2.6, respectively) were of a lesser magnitude than that seen for the VSG 221 negative GFP positive cells.

The rapid formation of DSBs at the I-*Sce*I restriction site was confirmed by Southern blot detection of the 9kb XhoI fragment (Figure 3.9a) using a probe specific to *PAC* (Figure 3.10e) (Glover et al., 2013a). In uninduced cell populations the fragment was present, but upon induction of I-*Sce*I expression, and the subsequent digestion of the VSG 221 ES, the fragment was lost. All three clones demonstrated inducible I-*Sce*I expression and efficient DSB formation within 24hours of doxycycline addition. Levels of detection of the 9kb fragment were compared to the parental VSG^{up} line prior to GFP transfection.

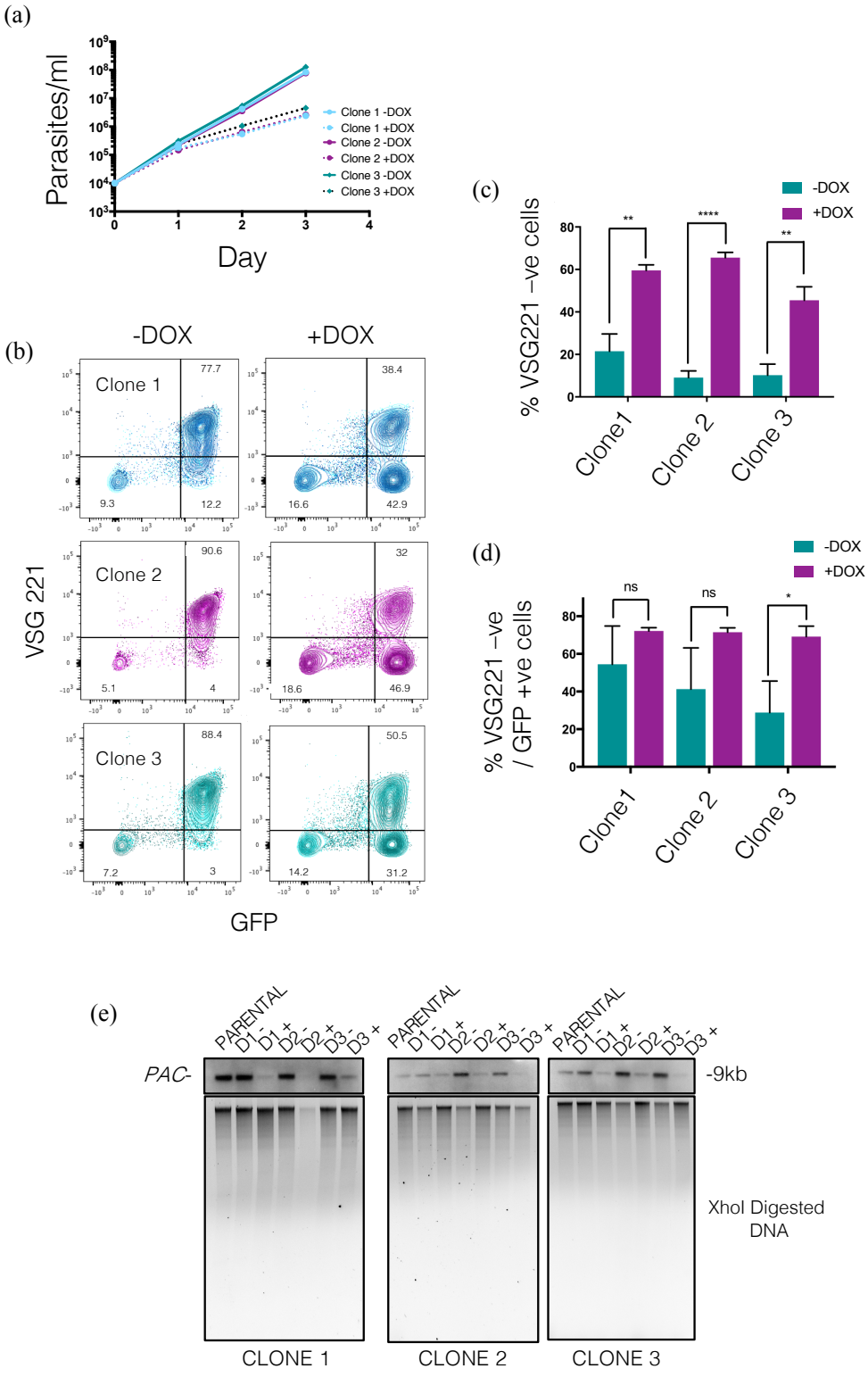


Figure 3.10 DNA recombination-mediated VSG switches were detected upon induction of I-*Sce*I expression in three VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones (a) Cumulative *in*

vitro growth of three uninduced (bold line) and induced (dotted line) VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones. Three replicates were performed for each condition and clone. Doxycycline was added on day 0 and replaced following addition of fresh HMI-9. Data represents the mean ± S.D., n=3 (where error bars were smaller than the symbols they were not drawn). **(b)** VSG 221 and GFP expression of cells isolated on day 3. A total of 10,000 events were captured. The plots show the data for all three triplicates superimposed onto each other. The mean proportions of cells in the quadrants are shown (n=3). VSG 221 expression was determined by α-VSG 221 staining. No antibody was required to detect GFP fluorescence. The VSG 221 positivity gate was determined using a VSG 221 positive control (the logarithmic population from which the assays were started), a VSG AnTat1.1 expressing negative control and a VSG 221 expressing secondary only control. The GFP positivity gate was determined using a GFP positive control (the logarithmic population from which the assays were started) and a cell line which had not been transfected with GFP for a negative control. **(c)** The percentage of cells which were classified as VSG 221 negative on day 3. Upon induction of I-*Sce*I expression, populations were significantly enriched for VSG 221 negative cells. Data represents the mean ± S.D. (n=3). ** = $p \leq 0.01$, **** = $p \leq 0.0001$, t-test. **(d)** The percentage of VSG 221 negative cells which were GFP positive on day 3. Data represents the mean ± S.D. (n=3). * = $p < 0.05$. **(e)** Southern blot detection of the 9kb XhoI DNA fragment with a *PAC* specific probe. Rapid loss of the 9kb fragment was observed following the induction of I-*Sce*I expression. Presence of the 9kb fragment was compared to the uninduced VSG^{up} parental stock (no GFP expression). The XhoI digested DNA is shown below as a loading control.

The FACS-based VSG switch assay was therefore able to both detect induced VSG switches and infer whether an ES switch or DNA recombination-based VSG switch had occurred. This allowed us to estimate the relative switch rate of a population and thus, by monitoring the proportion of the population consisting of VSG switched cells, we could deduce whether a given treatment influenced VSG switching. I next decided to test if absolute VSG switch frequency could be accurately calculated by the assay. Upon induction of DSBs upstream of the expressed VSG gene, Glover et al. (2013) estimated that the VSG switch rate increased 5000-fold to 5×10^{-2} switches/induced cell. Critically, however, this increase in VSG switch rate was compared to published estimates for laboratory-adapted

monomorphs, and not an untreated control. Therefore, the antigen switch rate of the uninduced VSG^{up} line was assumed to be ~1x10^{-5} switches/cell/generation, though this may not have been the case. The induced assays were each ran for one day longer than the respective uninduced clone's assay (Figure 3.11a). However, for each of the three clones, there was no statistically significant difference ($p > 0.05$, student's t-test) in the mean number of generations that the uninduced and induced assays were run for (26.45 ± 0.92 vs 26.38 ± 0.68, 27.97 ± 0.44 vs 27.63 ± 1.63 and 26.72 ± 0.44 vs 26.11 ± 0.13 generations, respectively). Each of the three uninduced clones were estimated to switch their expressed VSG at a greater rate than the published estimates for monomorphic trypanosomes. Clone 1 was estimated to switch its expressed VSG at a rate of 9.56x10^{-3} ± 0.004 switches/cell/generation, Clone 2 at 2.03x10^{-3} ± 0.001 switches/cell/generation and Clone 3 at 5.24x10^{-3} ± 0.002 switches/cell/generation (Figure 3.11b). Upon induction, it appeared from the data presented in Figure 3.11b that the number of VSG switches/cell/generation increased. When the number of VSG switches/cell before normalization to the number of generations was considered, however, this equated to 1 (0.999 ± 0.0006, 0.998 ± 0.003 and 1 ± 0.00006 switches/cell, respectively) for each of the three clones, therefore indicating that all cells were VSG 221 negative. The most likely explanation for the entire population of cells staining as VSG 221 negative by the end of the VSG switch assay, was that an I-*Sce*I induced VSG switch preceded the first parasite division and therefore the negative cells were not real 'switchers' but the progeny of a cell which had already undergone a gene conversion (GC) VSG switch. Glover et al. (2013) demonstrated that, in two strains with I-*Sce*I restriction sites in other portions of their VSG 221 expression sites, DSBs had begun to be processed by 6 hours post induction (data prior to 24 hours induction with the VSG^{up} strain was not presented). Since the three VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones divided ~ every 6-8 hours, a GC switch occurring prior to division was feasible. In the event a parasite division occurred before a DSB was processed, it is still unsurprising that the whole population would be VSG221 negative since Glover et al. (2013) reported that 100% of their induced cells underwent DSB-mediated VSG switching. In the current experimental format where populations are established from a single cell, these induced assays are therefore not informative, and were not followed up.

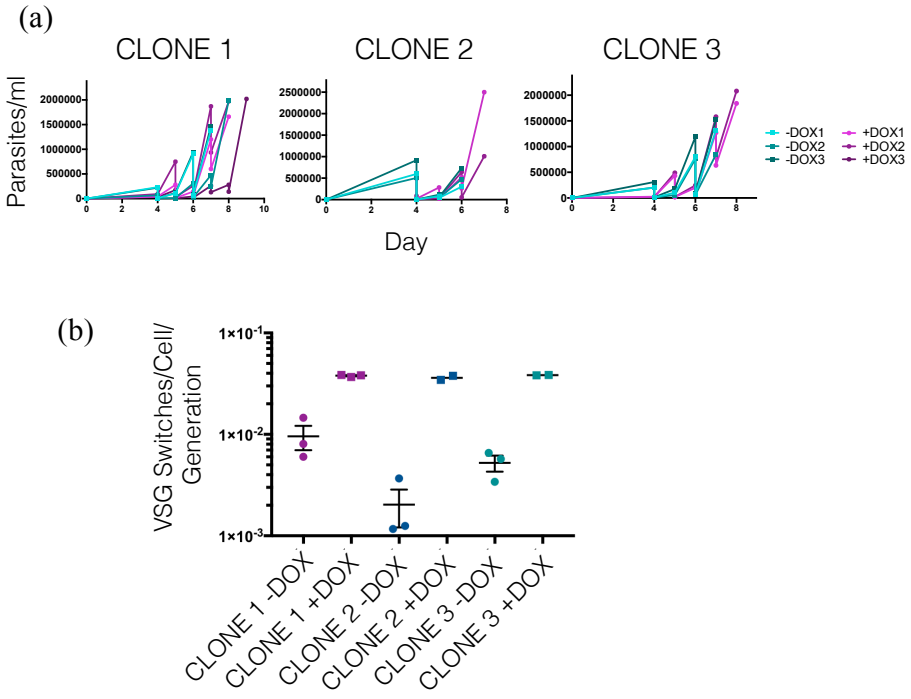


Figure 3.11 The FACS-based VSG switch assay could not be used to determine the VSG switch rate of induced VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones (a) VSG switch assay growth of VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones 1 (left), 2 (middle) and 3 (right). Uninduced populations are shaded in teal and induced in plum. Cells were induced on day 0. Assays were performed once in triplicate (n=3), except for the induced clone 2 and 3 replicates where n=2. **(b)** Calculated VSG switches/cell/generation for the uninduced (circular symbols) and induced (square symbols) populations of Clone 1 (plum), Clone 2 (blue) and Clone 3 (teal). Data represents the mean ± S.E.M (n=3, except for the induced clone 2 and 3 replicates where n=2).

Although the VSG switch rate of the uninduced VSG^{up} cells was never presented in the Glover et al. (2013) paper, the calculated VSG switches/cell/generation for the uninduced cells in Figure 3.11b still seemed very high in comparison to most other published VSG estimates for monomorphic Lister 427 lines (~1x10^{-7}-1x10^{-5} switches/cell/generation). It was therefore hypothesized that leaky expression of the I-*Sce*I enzyme could be the cause for the higher than

expected VSG switch rates. Although the Southern blot in Figure 3.10e confirmed that induction was inducible compared to the parental VSG^{up} stock, this was not proof that the parental stock itself was not inherently leaky. Therefore, the cell line depicted in Figure 3.12a was generated. A *T. brucei* Lister 427 cell line was transfected with the same I-*Sce*I restriction site, *PAC* gene and 5' and 3' UTRs targeting the VSG221 ES as were present in the original VSG^{up} strain, however it did not contain an inducible copy of the I-*Sce*I ORF and therefore no DSBs could be formed. Diagnostic PCRs confirmed that for two clones (1 and 3) the entire construct was present (primer set iv) and that it had integrated in the right orientation in the VSG 221 ES (primer sets i, ii and iii) (Figure 3.12b).

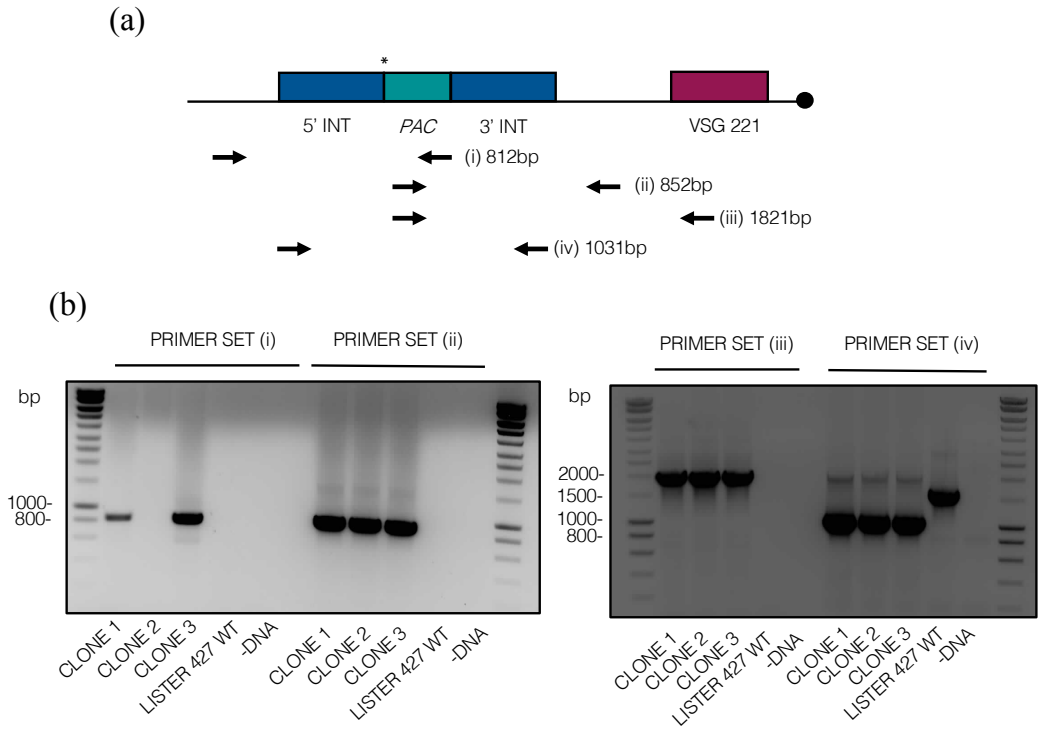


Figure 3.12 Generation of the VSG^{up} -I-*Sce*I cell line (a) An I-*Sce*I restriction site (*) and *PAC* gene were targeted to the VSG 221 ES, adjacent to the 70bp repeats. The cell line did not express an inducible I-*Sce*I meganuclease and therefore could not be 'leaky'. Diagnostic primer pairs and the desired product size are shown below. Primer set (i) tested for correct 5'end integration. Primer set (ii) tested for correct 3' end integration. Primer set (iii) tested that the construct had integrated in the VSG 221 ES. Primer set (iv) confirmed the whole construct was present. INT= Integration region. **(b)** Results of the diagnostic PCRs. Clones 1

and 3 were positive in all 4 PCRs. Wild-type Lister 427 DNA and -DNA samples served as controls.

The expression of the 9kb XhoI fragment in day 1 uninduced and induced VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones was compared to the expression levels in the VSG^{up} -I-SceI Clone 3 cell line by Southern blot (Figure 3.13). The expression level of the 9kb XhoI fragment was equivalent between the cell line which contained no I-*Sce*I ORF and the three uninduced VSG^{up} pLF12 -I-*Sce*I eGFP_BLA clones, thus suggesting that expression of I-*Sce*I was well controlled and therefore that leaky expression of I-*Sce*I, and consequently the generation of DSBs in the VSG 221 ES, was not responsible for the high VSG switch rates measured in Figure 3.11b.

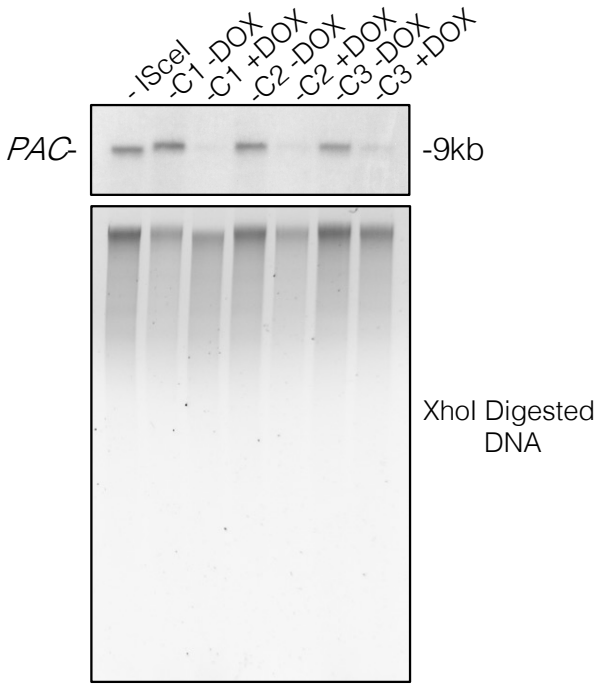


Figure 3.13 Leaky expression of the I-*Sce*I enzyme does not explain the high VSG switch rates calculated for the uninduced VSG^{up} pLF12 cells Southern blot detection of the 9kb XhoI DNA fragment with a *PAC* specific probe. Similar levels of *PAC* expression were obtained for the -I-*Sce*I control (far left) and the uninduced VSG^{up} pLF12 clones 1-3, therefore displaying that the enzyme's expression is inducible. The XhoI digested DNA is shown below as a loading control.

Using a monomorphic cell line with an inducible DSB site adjacent to the 70bp repeats in the active ES, we demonstrated that our FACS-based VSG switch assay was able to detect induced VSG switches and correctly identify whether a VSG switch was generated by a transcriptional switch or DNA rearrangement (Figure 3.10). The assay was therefore a useful tool to interrogate whether a given treatment (such as the induction of monomorphism) changed VSG switch rate. With regard to calculating the absolute (rather than relative) VSG switch frequency, our assay, admittedly, did have limitations. As discussed in Section 1.2.7, by calculating the proportion of a population which consisted of switched cells, one did not take into account the phenotypic or genetic diversity within this negatively staining population (Hovel-Miner et al., 2012) and nor was the potential difference in growth rates between variants accounted for. The VSG switch assay estimated that the uninduced (and monomorphic) VSG^{up} strain switched its expressed VSG at a rate of $5.24x10^{-3}$ – $9.56x10^{-3}$ switches/cell/generation, at least 100-fold greater than published estimated for laboratory-adapted trypanosomes (Figure 3.11). Generation of DSBs through leaky expression of the I-*Sce*I enzyme did not account for this high switch frequency. However, since the VSG switch rate of the uninduced parental line was not published, the ability of the VSG switch assay to accurately determine absolute VSG switch frequency cannot yet be determined definitively.

3.8 Inducibly monomorphic populations did not reduce their VSG switch rate upon induction

To investigate if induction of monomorphism was associated with a reduction in VSG switch rate, the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines were tested twice, in triplicate, by the *in vitro* VSG switch assay described in Section 3.6 (Figure 3.14a). Overall, the induced GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi switch assays ran for significantly longer than the uninduced assays (11 days vs 10.3±0.5 days, p=0.0250, Welch's t-test), suggesting that induction of RNAi may have caused a very slight growth defect. However, the mean number of generations completed during across the 6 uninduced and induced assays (26.8±0.4 vs 26.5±0.6,

respectively) were not significantly different ($p>0.05$, student's t-test). In the GFP^{ESpro}AnTat1.1^{ES} *NEK* and *DYRK* RNAi switch assays, there was no significant difference between the uninduced and induced conditions for either the mean number of days (9.5±0.5 vs 9 and 9.5±1.6 vs 8.4 ± 0.5, respectively) or the mean number of generations (26.9±0.6 vs 26.4±0.3 and 27.1±0.7 vs 27.3±1.2, respectively) that the assays were run for ($p>0.05$, student's t-test).

Pleomorphic trypanosomes are estimated to switch their expressed VSG at a rate of 10^{-4} to 10^{-2} switches/cell/generation (Turner and Barry, 1989; Turner, 1997) and Figure 3.9b shows that the calculated VSG switch rates for the uninduced, differentiation competent GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines were in agreement with these estimations ($6.97x10^{-4} \pm 0.0004$, $6.3x10^{-4} \pm 0.0002$ and $2.07x10^{-3} \pm 0.002$ switches/cell/generation, respectively). However, upon the induction of RNAi, and therefore the loss of differentiation competence, each of the three inducibly monomorphic trypanosome lines did not reduce their respective VSG switch rates to become equivalent to the estimated 10^{-7} to 10^{-5} VSG switches/cell/generation of laboratory adapted trypanosomes. In fact, there was a significant increase in switch frequency upon induction of RNAi against *HYP2* (to $2.51x10^{-3} \pm 0.0009$ switches/cell/generation $p=0.0008$, student's t-test). *NEK* knockdown caused a slight increase in VSG switching (to $7.34x10^{-4} \pm 0.0002$ switches/cell/generation) and *DYRK* knockdown decreased the VSG switch rate to $1.2x10^{-3} \pm 0.0007$ switches/cell/generation, however neither of these changes were significant ($p>0.05$, student's t-test).

The VSG switched cells were analysed for GFP positivity (which would indicate that they had switched their VSG by DNA recombination as opposed to a transcriptional switch) (Figure 3.9c). Differentiation-competent pleomorphic cells preferentially use DNA recombinatorial mechanisms to switch expressed VSGs (Robinson et al., 1999), whilst laboratory-adapted monomorphs have been shown to predominantly switch transcription between ESs instead (Liu et al., 1985; McCulloch et al., 1997). The VSG AnTat1.1 negative cells in the uninduced GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi populations appeared to be mostly GFP positive, whilst the VSG AnTat1.1 negative cells in the uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi

populations were roughly evenly split between the GFP positive and negative gates, however these results showed considerable variability (Figure 3.9c). Compared to the uninduced populations, there was a small decrease in the presence of GFP positive AnTat1.1 negative cells in the induced $GFP^{ESpro}AnTat1.1^{ES}$ *NEK* (52.7 ± 24.5% vs 49.75 ± 20.4%) and *DYRK* (84.63 ± 11.2% vs 70.56 ± 21.5%) RNAi populations following the induction of monomorphism, however these changes were not significant ($p>0.05$, student's t-test). The proportion of GFP positive VSG AnTat1.1 negative cells in the induced $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2* RNAi populations increased from 44.67 ± 25% to 71.68 ± 22.6%, however such was the variability in these experiments that this result was not significant ($p>0.05$, student's t-test).

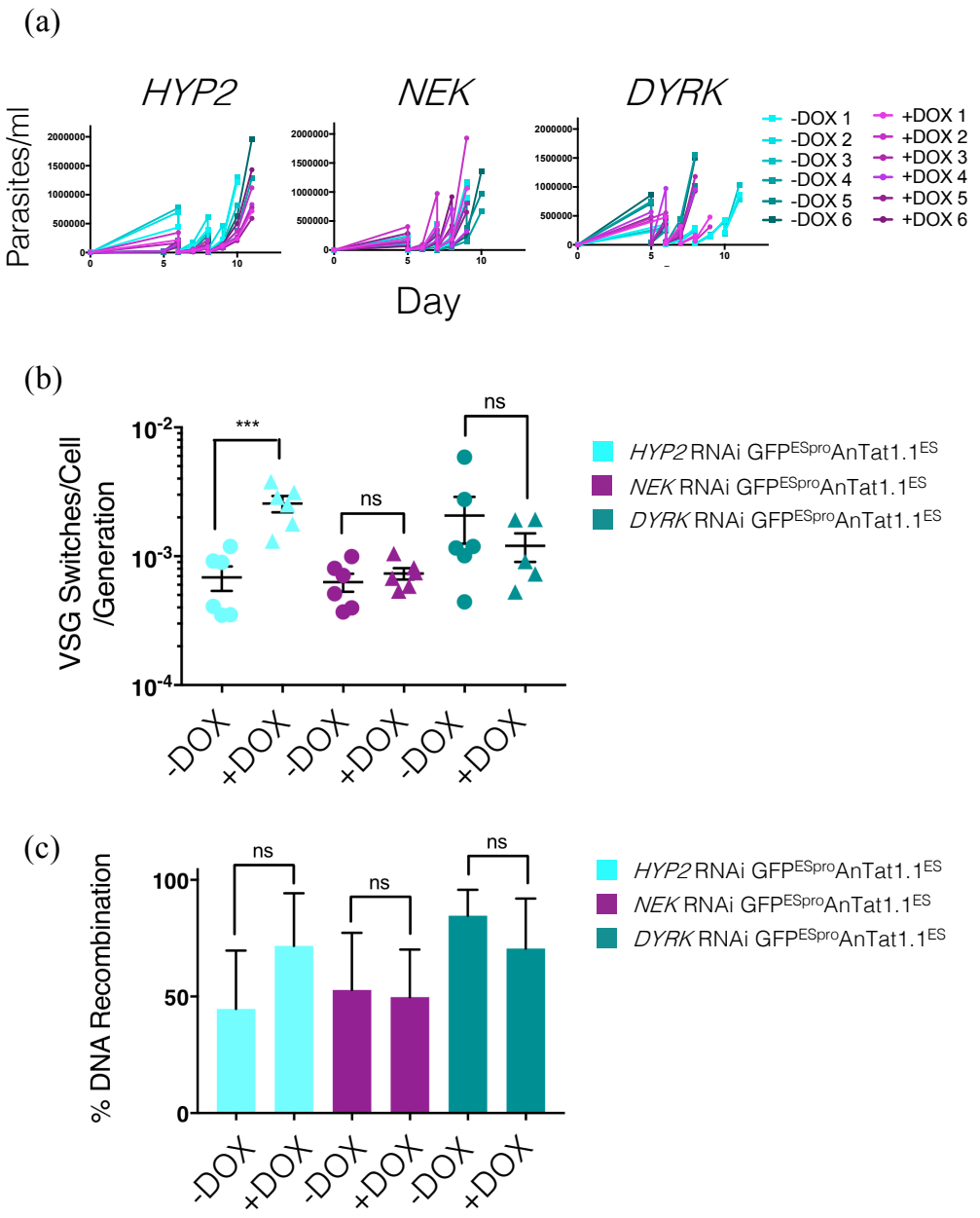


Figure 3.14 The induction of monomorphism did not significantly reduce VSG switch rate nor change the predominant mechanism of VSG switching (a) VSG switch assay growth of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* (left), *NEK* (middle) and *DYRK* (right) RNAi lines. Uninduced populations are shaded in teal and induced in plum. Assays were performed twice in triplicate (n=6), except for the induced *DYRK* RNAi experiment where one assay was performed in triplicate and one in duplicate (n=5). **(b)** The calculated VSG switches/cell/generation for the uninduced and induced GFP^{ESpro}AnTat1.1^{ES} *HYP2* (blue), *NEK* (plum) and *DYRK* (teal) RNAi lines. Upon the induction of RNAi against *NEK* and *DYRK*, there was no significant change in VSG switch rate compared to the uninduced

pleomorphs (p>0.05, student's t-test). VSG switching increased significantly upon the induction of RNAi against *HYP2* (p=0.0008, student's t-test). ***= p<0.001. **(c)** The percentage of switched GFP^{ESpro}AnTat1.1^{ES} *HYP2* (blue), *NEK* (plum) and *DYRK* (teal) RNAi cells which had switched their expressed VSG by DNA recombination (i.e. VSG AnTat1.1 negative/ GFP positive). There was no significant difference in the percentage of VSG switches elicited by DNA recombination upon the induction of RNAi against *HYP2*, *NEK* or *DYRK* (p>0.05, student's t-test).

The data collected from the *in vitro* FACS-based VSG switch assays showed that, upon the inducible knockdown of a component of the stumpy formation pathway (and therefore the loss of stumpy differentiation capacity) in three different RNAi lines, there was no significant reduction in VSG switch frequency, nor a significant change in the proportion of switches elicited by either DNA recombination or *in situ* switches of transcription (Figure 3.14). The results, therefore, strongly suggest that stumpy differentiation capacity and VSG switch frequency can be uncoupled. It is interesting to note that the GFP^{ESpro}AnTat1.1^{ES} HYP*2* and *NEK* RNAi lines, which displayed the lowest switch frequencies, also switched least by DNA recombination when compared to the faster switching GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi line. This is reminiscent of the differences in switch frequency and mechanism between pleomorphic (frequent VSG switching and preferentially by DNA recombination) and laboratory-adapted monomorphic (infrequent VSG switching and preferentially by switches of transcription) population. That the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi lines have a reduced VSG switch rate and frequency of DNA recombination-based VSG switching compared to the GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi cell line, yet still retain pleomorphism, supports the hypothesis that reduced VSG switch frequency and monomorphism can be independently selected for during laboratory passage, but at different rates. Without knowing the VSG switch rate of the *HYP2* or *NEK* RNAi lines at different periods of their laboratory passage prior to this study, however, this is difficult to demonstrate.

3.9 A monomorphic GFP^{ESpro} 221ES Lister 427 cell line switches expression of VSGs as frequently as pleomorphic AnTat1.1 90:13 cells

As part of the VSG switch assay validation experiments in Section 3.7, a monomorphic GFP^{ESpro} 221ES Lister 427 cell line from our lab was tested in the FACS-based VSG switch assay format using a VSG221 antibody (Figure 3.15). We found that our monomorphic cell line switched its expressed VSG at a mean rate of $5.99x10^{4} \pm 0.0004$ switches/cell/generation (Figure 3.15b, left) which is comparable to the rates of antigenic variation seen in the pleomorphic GFP^{ESpro} AnTat1.1^{ES} *HYP2* and *NEK* RNAi cell lines and less than two-fold slower than the fast switching *DYRK* RNAi cell line (Figure 3.14). In agreement with the literature, however, this monomorphic cell line did preferentially switch by *in situ* switches of transcription between expression sites (31.01 ± 15.6% *in situ* switches).

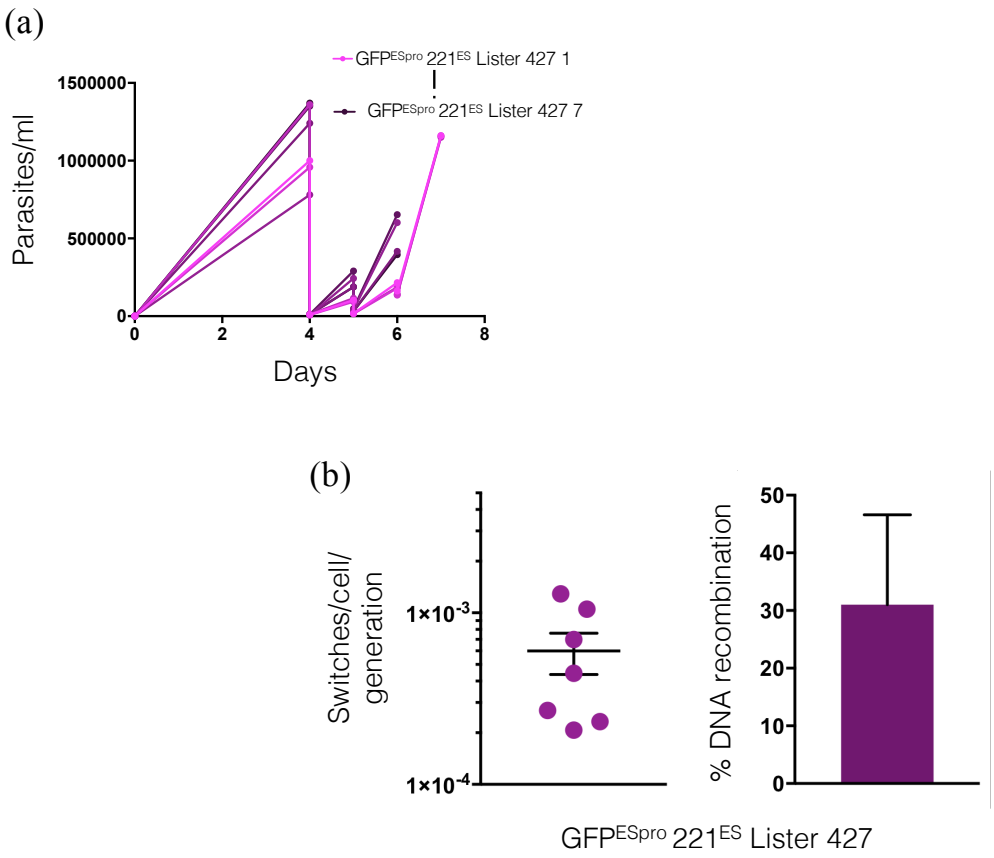


Figure 3.15 A GFP^{ESpro} 221ES Lister 427 cell line preferentially switched VSG expression by *in situ* switches of transcription at a high rate (a) VSG switch assay growth of the GFP^{ESpro}221ES Lister 427 replicates. Replicates 1-3 were performed as one assay and replicates 4-7 as another. The assays lasted for an average of 6.5 ± 0.5 days and the populations were fixed following a mean of 27.4 ± 0.9 generations. **(b) Left:** The calculated VSG switches/cell/generation for the 7 GFP^{ESpro}221ES Lister 427 replicates. The monomorphic populations switched expression of their VSG at a similar rate to pleomorphic cells. Data represents the mean ± S.E.M (n=7) **Right:** The percentage of switched GFP^{ESpro}221ES Lister 427 cells which had switched their expressed VSG by DNA recombination (i.e. VSG 221 negative/ GFP positive). The monomorphs preferentially used ES switching to switch VSG expression. Data represents the mean ± S.D (n=7).

The published VSG switch rates of the *T. brucei* Lister 427 cell line typically span 10^{-7} to 10^{-5} switches/cell/generation, and therefore our rates represent a ~10-1,000-fold increase on these. We would not expect integration of a GFP reporter

proximal to the ES promoter to induce VSG switches, since it has only been reported that disruption of the region directly upstream of the expressed VSG gene, the co-transposed region, causes an increase in VSG switching (Davies et al., 1997). To the best of our knowledge, since being in our hands, this clone has not been tested in any other VSG switch assay format.

3.10 Discussion

The cyclical passage of *T. brucei* through laboratory rodents selects against tsetse transmission potential and generates highly virulent populations of laboratory-adapted monomorphic trypanosomes. It has long been reported that these monomorphic lines of *T. brucei* switch their expressed VSG at a rate significantly lower rate than biologically relevant pleomorphic populations (Turner and Barry, 1989). Using the same cell line and method of measuring VSG switch frequency, Turner (1997) demonstrated that cells that had been subject to cyclical syringe passage switched expression of their VSG at a lower rate than cells that had recently been transmitted through a tsetse fly (albeit not consistently so). It has never been addressed, however, whether the loss of stumpy differentiation capacity that occurs upon laboratory-adaptation is directly associated with a concomitant reduction in VSG switch frequency, or whether the two processes, though both selected by multiple passage, are in fact independent.

In this Chapter, I have studied pleomorphic *T. brucei* cells lines which, upon the induction of RNAi against a particular gene in the stumpy formation pathway (either *HYP2*, *NEK* or *DYRK*), turned-off the capacity of the cell line to differentiate to transmission competent stumpy forms (Figure 3.5). This 'inducible monomorphism' model allowed us to probe VSG switching in the same cell lines where pleomorphism could either be switched 'on' or 'off', without long term passage. In our VSG switch assay, we have shown that the induction of monomorphism did not provoke a corresponding drop in VSG switch rate for any of the three RNAi cell lines (Figure 3.14). A GFP reporter integrated at a promoter-proximal position in the active VSG ES allowed us to also address whether the

populations were switching VSG by DNA recombination or *in situ* switches of transcription. It has been suggested that monomorphic strains preferentially use *in situ* switches of transcription (Liu et al., 1985; McCulloch et al., 1997); however, upon the induction of monomorphism we did not see a reduction in the mean percentage of switches that were mediated by DNA recombination. Our results therefore present evidence that *T. brucei* antigen switch frequency and stumpy differentiation capacity can be uncoupled.

In the VSG switch assays included in Figure 3.14, it was observed that the induction of RNAi against *HYP2* caused a significant 3.5-fold increase in VSG switch rate, from $6.97x10^{-4} \pm 0.0004$ to $2.5x10^{-3} \pm 0.0009$ switches/cell/generation. Therefore, *HYP2* could be involved in both the regulation of antigen switching and differentiation to stumpy forms. It must however be noted that we have been unable to repeat this VSG switching phenotype upon the induction of RNAi in further experiments and that VSG switch rate stays relatively static upon the knockdown of *HYP2*. We do not yet know the source for this experimental variability. If we were able to replicate the increase in VSG switch rate upon the induction of monomorphism, it would be interesting to perform single-cell RNAseq on induced cells to ask if the knock-down of *HYP2* was causing an increase in VSG switch rate or causing a deregulation of antigen expression. For example, the loss of monoallelic control could result in cells expressing two, or more, VSG surface coats, which would then be gated as a bona fide switcher in our FACS analysis.

Our FACS-based VSG switch assay was validated through the use of the VSG^{up} pLF12 -I-*Sce*I eGFP cell line. Upon the induction of I-*Sce*I expression in three VSG^{up} pLF12 -I-*Sce*I eGFP clones, our analysis indicated that the induced populations were significantly more enriched for VSG 221 negative cells compared to uninduced populations ($p<0.05$, student's t-test) and that 71 $\pm$ 3.4% of the induced switched populations were gated as GFP +ve (n=9), thus demonstrating that the FACS-assay could both detect VSG switches and DNA recombination. Within this assay format, however, our use of a promoter-proximal eGFP ORF as a marker for ES transcriptional activity does have limitations. Principally, whilst promoter-proximal GFP mRNA levels are rapidly attenuated to 20% of the original level

within 24 hours of ES shut down (Zimmermann et al., 2017), GFP protein has been shown to have a half-life of ~80 hours in *L. infantum* promastigotes (Kamau et al., 2001) and ~26 hours in mammalian cells (Corish and Tyler-Smith, 1999). Aresta-Branco et al. (2016) demonstrated by FACS that the fluorescence emitted by trypanosomes expressing a promoter proximal GFP ORF matched that of the negative control 120 hours after transcriptional shut down of the active VSG ES (though earlier timepoints were not tested). This suggests that switched cells may still express residual GFP, albeit at decreasing levels, for up to 5 days after a transcriptional switch. Nevertheless, this is not the first study to use GFP fluorescence encoded from a promoter-proximal GFP ORF as a marker of *T. brucei* BES activity. Both Aitcheson et al. (2005) and Devlin et al. (2016) used GFP negativity, as determined by IFA analysis, for classifying cells which had switched transcription of ESs. Since flow cytometry can detect changes in fluorescence to a far better degree than the human eye, our method of quantitating GFP fluorescence is more accurate than the use of microscopy alone. We have tried to circumvent the underestimation of ES switching by positioning the gate defining GFP negativity as close to the positive control peak as possible without introducing false negatives (as discussed in Section 3.6.2). Nonetheless, an improvement to our method could involve the incorporation of a reporter protein which is more rapidly turned-over following transcriptional shut-down. Fusion of the C-terminus PEST sequence of mouse ornithine decarboxylase (MODC) to eGFP has been shown to reduce the half-life of the protein to 2 hours in mammalian cells (Li et al., 1998). Of course, if using such a method it would have to be established that the protein was stable enough to retain fluorescence following fixation for FACS so that false negatives would not be introduced into the analysis.

We must acknowledge, that in these *in vitro* VSG switch assay experiments, we did not validate the knockdown of the *HYP2*, *NEK* or *DYRK* transcripts at the end of the switch assays (in order to preserve material for FACS analysis), nor demonstrate the loss of differentiation capacity in the induced population compared to the uninduced. Regarding the transcript knockdown, growth assays (which were used for VSG diversity studies) demonstrated efficient knockdown of *HYP2*, *NEK* and *DYRK* by qRT-PCR using identical RNAi induction conditions as were used in

this Chapter (see Chapter 4, Section 4.2.2). Furthermore, consistent growth phenotypes upon the induction of RNAi were observed, furthermore suggesting that the induction conditions were working. Upon induction of *HYP2* RNAi, the cells consistently proliferated at a lower rate, with the induced assays lasting significantly longer than the uninduced VSG switch assays (Figure 3.14a). In contrast, the induced *DYRK* RNAi cells were consistently observed to proliferate marginally quicker than the uninduced cells with the induced assays lasting a mean of 8.4 ± 0.5days compared to 9.5 ± 1.6 days for the uninduced cells (this difference was not significant however). Regardless, future VSG switch assays should make sure to extract RNA from the final population to confirm the respective transcript knockdown via qRT-PCR. In this assay's format, it is not possible to directly compare the differentiation capacity of the uninduced and induced populations via, for e.g., attempting to stimulate *in vitro* differentiation of the populations to the insect procyclic form by cold shock and treatment with *cis*-aconitate. This is because the uninduced population is maintained at densities which would not provoke a density-dependent differentiation response to stumpy forms and therefore would not permit further development through the life-cycle. Uninduced populations are maintained at these densities since differentiation to the stumpy form would essentially decrease the VSG switch rate of the population as the stumpy cells are committed to cell cycle arrest and cannot generate new variants.

In our assay's current format, VSG AnTat1.1 negative/GFP negative cells are presumed to have changed their expressed VSG via an *in situ* switch of transcription. It must be acknowledged however, that if a gene conversion event were to occur anywhere in the 463bp between the ES promoter and the eGFP ORF, the same FACS profile would be produced. Duplication of an entire BES is rarer than ES switching, but it does happen- in a *T. brucei* Lister 427 strain, Hertz-Fowler et al. (2008) found that 3/28 switched clones had replaced their entire BES via duplicative transposition of the new BES (Hertz-Fowler et al., 2008). To the best of our knowledge, no VSG switch assay using traditional PCR, FACS or IFA methods has found a way to distinguish *in situ* switches (with deletion of the old ES) from duplications of an entire BES. In the pleomorphic $GFP^{ESpro}AnTat1.1^{ES}$ *DYRK* RNAi populations, we found 84.63 ± 11.2% of switches were mediated by DNA recombination, (Figure

3.14). An analysis of the first two relapse peaks of a single *T. brucei* EATRO 797 rabbit infection isolated 88 switched clones which expressed 11 different VATs in total Robinson et al. (1999). In this pleomorphic cell line 11/13 VSGs (84.6%) were activated by duplicative gene conversion. With these results in mind, it seems unlikely that we are significantly underestimating the proportion of switches that occur by DNA recombination. Although the uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi lines were demonstrated to switch less by DNA recombination, this may have biological significance (as discussed in Section 3.8).

In our VSG switch assay, we found that a monomorphic GFP^{ESpro} 221ES Lister 427 cell line switched expression of VSGs at a rate comparable to the pleomorphic RNAi lines tested in Section 3.8. Although most published estimates of VSG switch frequency in monomorphic *T. brucei* Lister 427 strains are in the range of 10^{-7}-10^{-5} switches/cell/generation, there is evidence of monomorphic trypanosomes switching their expressed VSG more frequently. Aitcheson et al. (2005) measured the VSG switch rate of an induced *T. brucei* Lister 427 90:13 cell line expressing an RNAi construct targeting VSG221 and a promoter proximal eGFP ORF as approximately $1x10^{-4}$ switches/cell/generation in their *in vitro* switch assays. Fluctuation analysis confirmed that the observed VSG switches were occurring spontaneously and not as a result of the induction of VSG 221 RNAi (Luria and Delbruck, 1943). Aitcheson et al. (2005) reasoned that the experimental procedure, in particular the method of negative selection against VSG 221 expressing parasites, could be the reason that more switched cells are detected. They hypothesised that negative selection using VSG 221 RNAi perhaps allowed the recovery of cells which had recently switched VSG expression (and therefore contained the transcripts of the old and new VSGs). This contrasts with another *in vitro* negative selection method using TK and Ganciclovir (that generated switch frequencies in the range of $1\text{-}3x10^{-5}$ switches/cell/generation (Cross et al., 1998) that only recovered early switchers where the expressed TK transcripts had dropped to very low levels (Aitcheson et al., 2005). Using a similar method of analysis as it utilised for our VSG switch assays, Briggs et al. (2018) also detected rates of VSG switching in wild type *T. brucei* Lister 427 cells that were higher than would normally be expected. At four timepoints over the course of ~45 generations *in vitro*, cells were fixed and analysed

for VSG expression by IFA using α-VSG 221 antiserum. Without any negative selection against VSG 221, ~1-2 out of every 1000 wild type cells were detected that had switched expression away from VSG 221 (Briggs et al., 2018).

Why then might we, and others, measure relatively high VSG switch rates for monomorphic cell lines? We must firstly consider that in these studies, and indeed most other VSG switch studies, although the genetic background being used is the same (i.e. *T. brucei* Lister 427), the cells will have been handled differently between laboratories. For example, some clones may have been subject to years of high density passage whilst others have been kept at a (relatively) low passage number. As a result of these disparities, 'laboratory specific' adaptations may occur in given cell lines and influence the biology of the parasite. As it is not uncommon to see different growth rates between different clones of the same line, perhaps we could also expect to see differences in VSG switch rates. Although we know that long-term passage in the absence of tsetse transmission causes monomorphism, we do not know the specific mechanism behind this loss in response to SIF. For instance, if monomorphism has arisen multiple times but by different means in each laboratory then the 'umbrella term' monomorph could actually comprise many different phenotypes, including high or low VSG switch frequency.

Although phenotypic differences between cell lines may account for why we measured a switch rate of $5.99 \times 10^{-4} \pm 0.0004$ switches/cell/generation in our monomorphic GFP^{ESpro} 221^{ES} Lister 427 cell line, it does not explain our finding in Figure 3.11 where we measured the mean VSG switch rate of three uninduced VSG^{up} pLF12 -I-*Sce*I eGFP clones as $5.61 \times 10^{-3} \pm 0.004$ switches/cell/generation. The parental uninduced VSG^{up} line was published to switch VSG expression at a rate of $\sim 1 \times 10^{-5}$ switches/cell (however, this was just an estimation of VSG switch frequency compared to published values for wild type monomorphic cells (Glover et al., 2013a)). As discussed in Section 1.2.7, there is a wide range of *in vivo* and *in vitro* techniques that are used to measure VSG switch rate. An explanation for why we (and others) have estimated monomorphic *T. brucei* Lister 427 cell lines to switch at relatively high frequency could lie in the chosen methods of analysis. As previously mentioned, Aitcheson et al. (2005) hypothesised that certain methods of negative

selection could remove cells which have only very recently switched expression from one VSG to another. If this is true, then it is less surprising that in studies such as our own and Briggs et al. (2018), where no negative selection against the initiating VSG was performed, the highest VSG switch rates are generated. With *in vivo* experiments, or with *in vitro* experiments utilising VSG-specific antiserum, cross reacting antibodies may clear newly generated variants whose epitopes are sufficiently similar to that of the VSG under negative selection. Since the number of surviving clones is incorporated into the calculation of VSG switch rate, as shown in Section 1.2.7, this would result in an underestimation of the VSG switches/cell/generation. As mentioned above, HSV-TK/Ganciclovir negative selection could result in an underestimation of switch frequency when newly switched variants still have enough TK to render them sensitive to Ganciclovir treatment. Similarly, MACS depletion of the initiating VSG could also underestimate the VSG switch frequency if a newly switched variant still retained enough of the initial VSG on its surface to bind the magnetic column. A negative selection step against the initiating VSG, whether by immunological or biochemical means, could therefore lead to an underestimation of VSG switch rate. However, that is not to say that the approach used in our study, as well as those adopted by Aitcheson et al. (2005) and Briggs et al. (2018), have not overestimated VSG switch rate of the monomorphic *T. brucei* Lister 427 line. Suboptimal fixation and α-VSG 221 staining conditions could influence the FACS or IFA interpretations (though such is the consistency of our observations that this seems unlikely), whilst Aitcheson et al. (2005) suggested that undetectable levels of RNAi 'leakiness' could also explain the relative increase in VSG switch rate.

As discussed in Section 3.6.2, the gate defining VSG AnTat1.1 or VSG 221 positivity was positioned directly to the left of the positive control to account for the time taken for cells to fully replace their VSG surface coat (Figure 3.8b). It was argued that placing the gate to the right of the secondary antibody only and negative controls would wrongly exclude cells from the analysis which had undergone a VSG switch, but had yet to completely replace their surface coats with the new VSG. Placing the gate so close to the positive control, however, could introduce false

negatives into the analysis. For instance, if there was suboptimal antibody staining then some cells may wrongly appear VSG negative or if cells which had recently divided and did not yet possess a full VSG AnTat1.1 or VSG 221 complement on their surface were present these too could be wrongly classified as VSG AnTat1.1/221 negative. If the gating strategy was introducing too many false VSG 221 negative cells this could account for the unusually high VSG switch rates measured in the monomorphic *T. brucei* Lister 427 lines. To ask if the FACS gating strategy was responsible for the high rates of VSG switching observed in the *T. brucei* Lister 427 lines, the data presented in Figure 3.15 was reanalysed by placing the gate defining VSG221 positivity directly to the right of the negative control (Figure 3.16). When the VSG 221 gate was moved, the mean VSG switches/cell/generation dropped only very marginally from $5.99x10^{-4} \pm 0.0004$ switches/cell/generation to $4.31x10^{-4} \pm 0.0003$ switches/cell/generation ($p>0.05$, Welches t-test). At $4.31x10^{-4} \pm 0.0003$ switches/cell/generation, this change in gating strategy still did not reduce to calculated VSG switch rate of the monomorphic population to the published VSG switch rates of between $1x10^{-7}$ and $1x10^{-5}$ switches/cell/generation. Therefore, whilst it must be acknowledged that the gating strategy adopted in this study could incorporate false negatives, the gating strategy is not the sole reason for the higher than expected VSG switch rates calculated for the monomorphic $GFP^{ESpro}VSG221^{ES}$ Lister 427 cell lines.

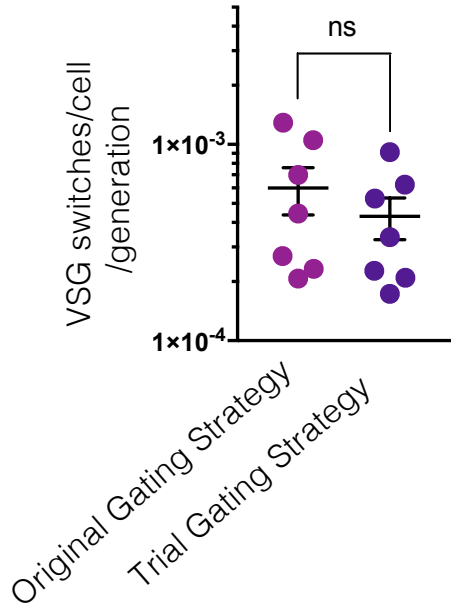


Figure 3.16 The FACS VSG gating strategy is not responsible for the calculation of high rates of VSG switching in the monomorphic *T. brucei* GFP^{ESpro}VSG221^{ES} Lister 427 strain The calculated VSG switches/cell/generation for the 7 GFP^{ESpro}221ES Lister 427 replicates when the original gating strategy was used (left) and when the gate determining VSG positivity was moved to the right of the negative control (right). When using the trial VSG gating strategy, the VSG switch rate dropped approximately 1.4-fold, however the change was not significant ($p>0.05$, Welches t-test) The data represents the mean ± S.E.M (n=7)

Together then, the use of laboratory specific stocks of Lister 427 cells and the application of different experimental techniques could explain why some studies, including ours, calculate the VSG switch rate of monomorphic *T. brucei* Lister 427 cells as ~10^{-3} to 10^{-4} switches/cell/generation rather than the more commonly quoted 10^{-7}-10^{-5} switches/cell/generation. If different cell stocks and switch assays can produce VSG switch rates differing by 1,000-10,000-fold, then this has huge implications when calculating absolute VSG switch frequency for any trypanosome cell line. In the absence of a gold standard of VSG switch assay, how do we decide what the 'real' VSG switch rate of our cell lines are? In the context of studies such as ours then, maybe the more useful question is not 'What is the absolute VSG switch

frequency?', but rather 'How does the relative VSG switch rate change within my experimental design?'.

In summary, we have used an 'inducible monomorphism' model to investigate if VSG switch frequency and slender to stumpy differentiation capacity are directly linked. Using our *in vitro* FACS-based VSG switch assay, we have demonstrated that the induction of monomorphism did not reduce the VSG switch rate nor change the preferred mechanism of VSG switching. In doing so, we have presented the first evidence that slender to stumpy differentiation capacity and antigen switch frequency can be uncoupled. We have also shown that the experimental design and use of laboratory-specific cell stocks can change the estimated VSG switch rate for a given genetic background by up to 1,000-fold. Therefore, in studies such as ours investigating the effect of a given treatment on VSG switch rate, a comparison of the relative VSG switch rate within a given experimental set-up is more informative than a comparison to the published absolute VSG switch rates in other studies.

4 : DETERMINING THE EXPRESSED VSG DIVERSITY OF PLEOMORPHIC AND INDUCIBLY MONOMORPHIC TRYPANOSOMES BY VSGseq

4.1 Introduction

The inducible knockdown of a gene involved in the stumpy formation pathway, and the subsequent loss of slender to stumpy differentiation capacity, did not elicit a reduction in VSG switch frequency compared to the parental pleomorphs (Figure 3.9). This suggests that antigen switch frequency and life cycle competency can be uncoupled, contrary to the common conception that monomorphic populations switch their expressed VSG at rates ~100-1000-fold less than pleomorphic populations. In Chapter 3, the VSG switch rates of the pleomorphic and inducibly monomorphic populations were calculated through *in vitro* FACS-based VSG switch assays, which are conceptually similar to other published methods. Following a short period of *in vitro* growth, VSG switched cells were quantified (either following an enrichment step or not, as is the case with the assays presented in Chapter 3) and the ratio of these compared to the total population analysed used to estimate VSG switches/cell (Boothroyd et al., 2009; Devlin et al., 2016; Hertz-Fowler et al., 2008; Horn and Cross, 1997; Hovel-Miner et al., 2016; Hovel-Miner et al., 2012; Kim and Cross, 2010; Schulz et al., 2016).

Recent advances in technology now allow us to assess VSG switching at a much deeper level by analysing the expressed VSG diversity within populations using Illumina deep sequencing (McCulloch and Field, 2015; Mugnier et al., 2015). In particular, VSGseq is a targeted RNAseq approach developed and validated by Mugnier et al. (2015). Exploiting the conserved 5' spliced leader and 3' 14mer sequences at the end of all mature VSG mRNAs, VSG cDNA is specifically amplified from total RNA. The VSG cDNA is sequenced and then VSG sequences assembled *de novo* (Mugnier et al., 2015). In this Chapter, this approach was used to quantitively examine the expressed VSG diversity in populations of pleomorphic and inducibly monomorphic cells to corroborate the VSG switch assay results from Chapter 3.

4.2 A workflow to determine the expressed VSG diversity incorporating MACS and VSGseq

For VSGseq analysis, the same GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi cell lines used in Chapter 3 would be studied when uninduced, and therefore differentiation competent, or induced, and so incapable of differentiation to stumpy forms. For consistency, the growth assays prior to RNA extraction were again be performed *in vitro*, scaling up culture volumes rather than diluting them, to prevent bottlenecking (see Section 4.2.1). Through discussions with Monica Mugnier (a collaborator in the study) the concern was raised that rare or minor VSG switch variants may not be detected due to the high prevalence of VSG AnTat1.1 expressors in the populations. Therefore, to increase the depth of analysis it was decided to deplete the populations of VSG AnTat1.1 expressors by MACS prior to the isolation of parasite RNA from the remaining switched cells (Boothroyd et al., 2009; Schulz et al., 2016). We were worried, however, that a MACS-enrichment step could introduce bias into the analysis of expressed VSG diversity or that MACS enrichment may not be equivalent between samples. Therefore, it was decided to collect a sample of total population RNA prior to MACS enrichment (this will be referred to as the "pre-MACS" sample), as well as isolating RNA from the VSG switch-enriched population after MACS selection (this will be referred to as the "post-MACS" sample). Analysis of the pre-MACS dataset would allow us to take an unbiased approach to quantitate overall VSG expression diversity, whilst the post-MACS datasets would instead allow us to examine in depth which VSGs were being activated and whether some VSG genes (or cohorts of genes) were preferentially activated by pleomorphs or inducible monomorphs.

The above considerations were consolidated into the workflow presented in Figure 4.1. At each step of the workflow prior to VSGseq, FACS and/or IFA samples were taken to either validate VSG expression or to confirm that both the uninduced and induced populations were proliferative. Triplicate pre- and post-MACS RNA samples for each uninduced and induced GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi population were sequenced (36 RNA samples in total).

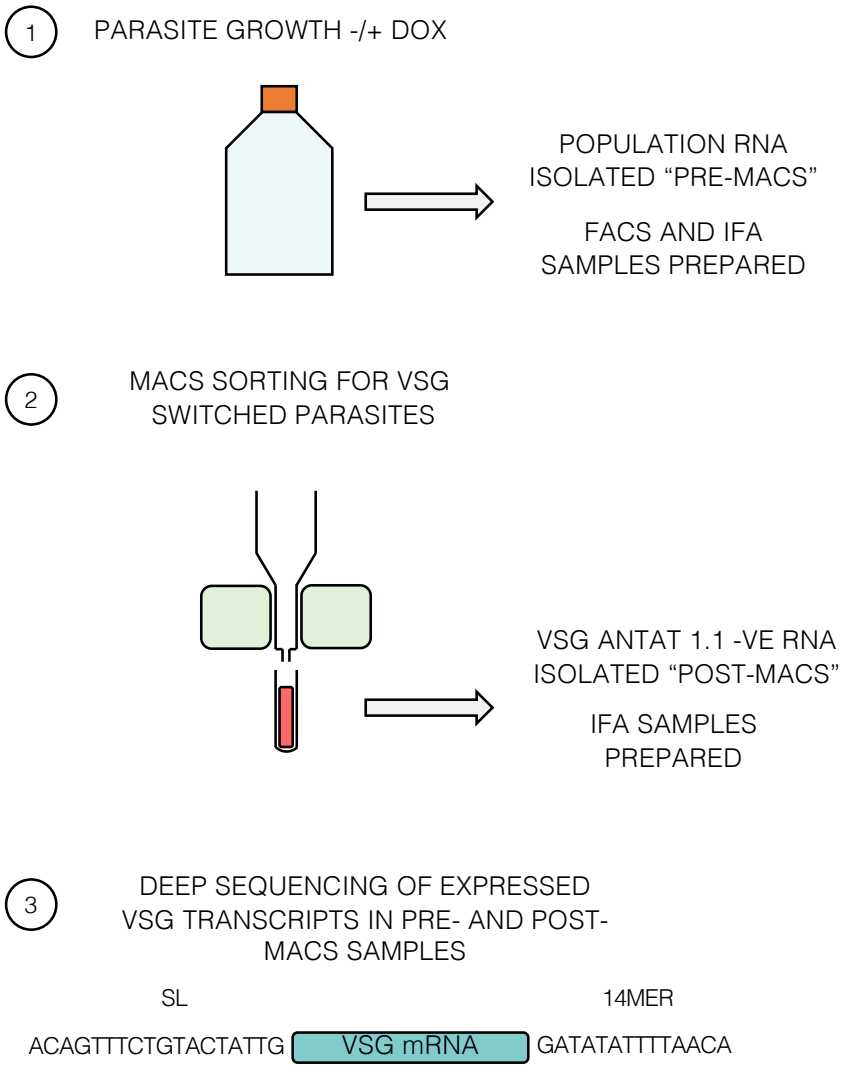


Figure 4.1 Workflow to determine expressed VSG diversity *in vitro* Following a period of *in vitro* growth in non-induced or induced conditions (Step 1), a fraction of the total parasite population was used to extract total RNA. This is referred to as the "pre-MACS" RNA sample and provided an unbiased analysis of overall VSG expression diversity within the population. At this stage a small number of parasites were further isolated and fixed for FACS analysis of VSG expression (to confirm the population originated from a VSG AnTat1.1 expressing cell) and for IFA analysis of cell cycle and PAD1 expression (to confirm the population was PAD1 negative and so slender and replicative). The remainder of the population was enriched for VSG AnTat1.1 negative cells by MACS. Cells were incubated with magnetic beads coupled to the α-VSG AnTat1.1 antibody before being

passed through a magnetic column (see Section 4.2.1) (Step 2). Following the MACS step, a small number of parasites were removed from the 'VSG AnTat1.1 negative' population for IFA analysis of VSG AnTat1.1 expression in order to validate that the enrichment step had worked. The remainder of the enriched population was used to extract RNA. This is the "post-MACS" RNA sample, used to determine if specific VSG genes or gene cohorts were preferentially expressed by the pleomorphic or inducibly monomorphic cells. This process was repeated in triplicate for all the induced and uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi populations, generating 36 RNA samples. The expressed VSG diversity in these 36 samples is assessed by VSGseq (Step 3) where VSG cDNA is specifically amplified from the SL and 14mer sequences, sequenced and assembled *de novo*.

4.2.1 MACS was effective for depleting a population of VSG AnTat1.1 expressing cells

Minor VSG switch variants (i.e. those which form $\leq$ 1% of the total population as defined by Mugnier et al. (2015)) could be lost to the background sequencing noise when the population is dominated by a single VSG (which in the case of the *in vitro* GFP^{ESpro}AnTat1.1^{ES} RNAi populations is VSG AnTat1.1). Mugnier et al. (2015) demonstrated that VSGseq could reliably detect VSG variants comprising 0.01% of the population, and accurately quantify a variant's presence within the population when the variant formed $\geq$ 0.1% of the population (Mugnier et al., 2015). To ensure the detection of switched variants, populations of cells were depleted of VSG AnTat1.1 expressing cells by MACS (Step 2, Figure 4.1) after the *in vitro* growth phase and the isolation of total population RNA (Step 1, Figure 4.1). Briefly, the population was incubated with a rabbit α-VSG AnTat1.1 antibody followed by α-rabbit magnetic beads such that, when passed through a magnetic column, the VSG AnTat1.1 expressing cells bind, whilst switched, antigenically distinct cells are collected in the flow-through (Boothroyd et al., 2009; Schulz et al., 2016). To prevent internalisation of the α-VSG AnTat1.1 antibody (the cells are not fixed prior to the procedure) all incubation, centrifugation and column steps were performed in a cold room at 4 ºC or on ice (Engstler et al., 2007; Schulz et al., 2016).

To generate switch events, populations of cell lines where RNAi targeted genes involved in stumpy formation (±induction) were expanded in non-selective medium from a single VSG AnTat1.1 and GFP positive cell as was performed in Chapter 3. The populations were maintained at $\leq 1x10^6$ cells/ml by the addition of HMI-9 medium to dilute the cultures, rather than remove cells. This strategy reduced bottle-necking during the workflow (Figure 4.2a). From the final population, RNA, FACS and IFA samples were required, in addition to sufficient cells for MACS. This required an additional scale-up step, bringing the final culture volume to 125ml and the total assay duration to 12 days (Figure 4.2a), thereby generating $1.25x10^8$ cells assuming the culture density was $1x10^6$ cells/ml. Extending the growth phase of the assay further, whilst maintaining cells in logarithmic growth, was not feasible due to the culture volumes required.

The capacity of the MACS columns used for the depletion of VSG AnTat1.1 expressing trypanosomes (LD Columns, Miltenyi Biotec) was $1x10^8$ magnetically labelled cells from a total population of $5x10^8$ cells. The FACS-based VSG switch assays presented in Chapter 3 demonstrated that, following 8-11 days growth in non-selective medium, ~0.01% of the population were VSG AnTat1.1 negative. Therefore, since the majority of the population would still be expressing VSG AnTat1.1 by day 12, it was decided that no more than $1x10^8$ cells should be loaded onto a MACS column. From the remaining ~$2.5x10^7$ cells in the flask, two aliquots of $1x10^6$ cells were each fixed for the IFA and FACS analysis and RNA extracted from the remaining ~$2.3x10^7$ cells (pre-MACS RNA). Thereafter, the Schulz et al. (2016) MACS protocol for the depletion of specific VSG expressors from an *in vitro* population was performed, as described in Section 2.17.2, using an α-VSG AnTat1.1 primary antibody. From the final column flow-through, 1ml was washed and fixed for IFA analysis of VSG expression and the remainder used for the post-MACS RNA extraction.

Using the above *in vitro* growth and VSG AnTat1.1 depletion conditions, 3 trial VSGseq workflows were performed using uninduced pLF12GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cells to validate the protocol. Two trials followed the full protocol through to post-MACS RNA extraction, whilst the third was used solely to validate

the effectiveness of the MACS depletion of VSG AnTat1.1 expressors from the population. To preserve as much material as possible for the post-MACS RNA extractions, FACS samples for assessing MACS efficiency were not removed from the column flow-through in the first two trials or the final sequencing experiments. The FACS profiles generated from the population before and after VSG AnTat1.1 depletion during the third trial workflow are shown in Figure 4.2b. Following 12 days *in vitro* growth, 87.7% of the unsorted pLF12 $GFP^{ESpro}AnTat1.1^{ES}$ population remained VSG AnTat1.1 positive (Figure 4.2b, left panel). Following MACS VSG AnTat1.1 depletion, the flow-through population was 43.8% VSG AnTat1.1 negative (Figure 4.2b, middle panel) demonstrating that the MACS step was effective at enriching for VSG AnTat1.1 negative cells, however that some positive cells were still present in the column flow-through. Although, for this single trial, 91.7% of cells gated as VSG AnTat1.1 negative in the unsorted population were GFP negative, 77.6% of the VSG AnTat1.1 negative cells in the column flow through were gated as GFP positive. As explained in Section 3.6.2, the gate defining VSG AnTat1.1 positivity was placed directly to the left of the VSG AnTat1.1 positive control peak. If the multiple steps of the MACS protocol were causing VSG to be shed from VSG AnTat1.1 positive and GFP negative cells, this could explain why a higher proportion of GFP positive cells were quantitated in the column flow-through. This hypothesis is supported by the observation that 8.84% of the column-bound population gated as VSG AnTat1.1 negative. Thus, VSG protein may have been shed during the multiple centrifugation steps or forceful plunging of bound cells from the column, or α-VSG AnTat1.1 antibody/magnetic beads used for the MACS step may have blocked binding of the α-VSG AnTat1.1 antibody used for the FACS analysis. Regardless of the reason(s) for this discrepancy, the MACS step of the VSGseq workflow was demonstrated to be efficient at depletion of VSG AnTat1.1 positive cells from the population, with the flow-through population being 43.8% VSG AnTat1.1 negative.

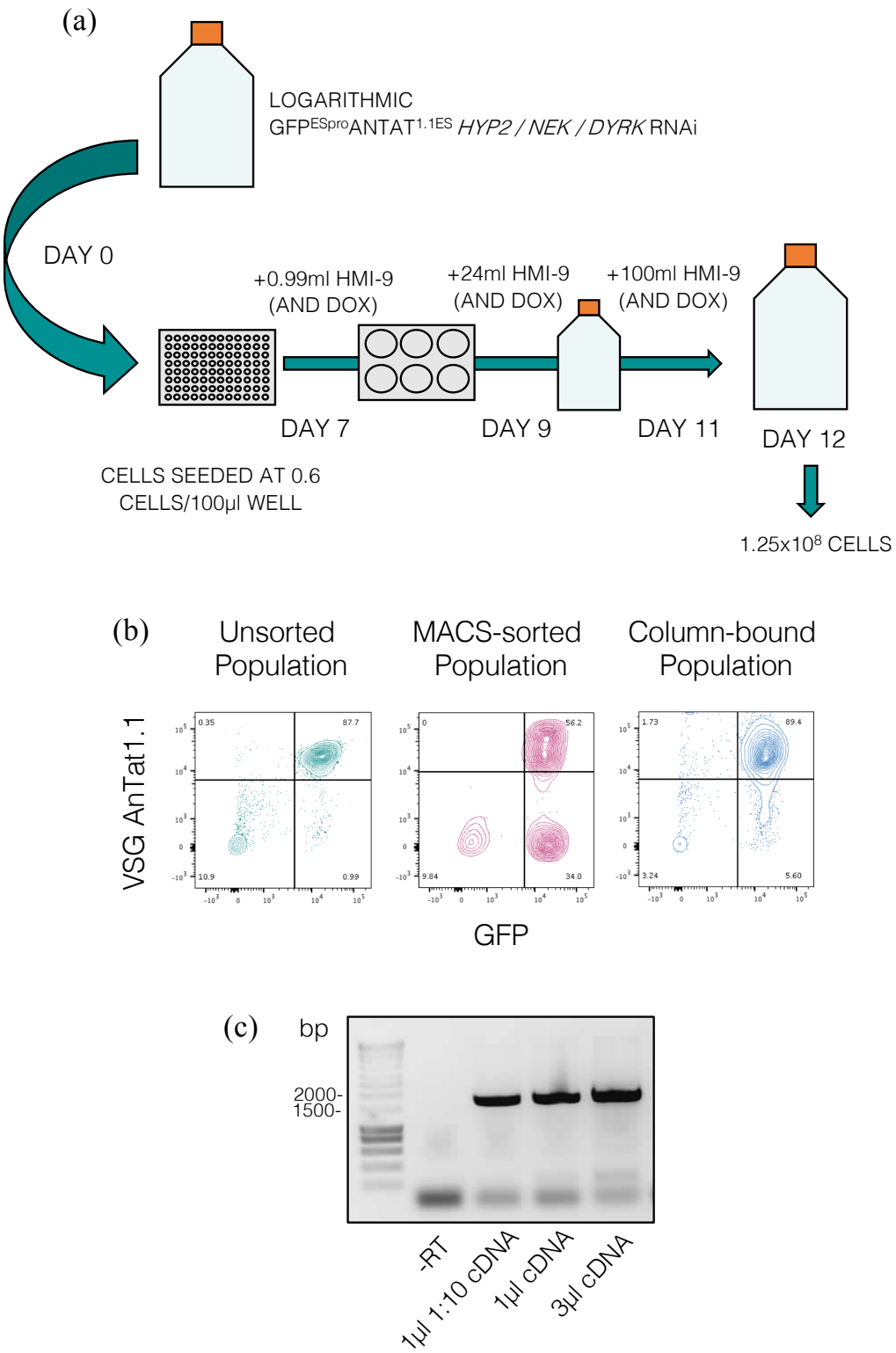


Figure 4.2 Validation of the proposed VSGseq workflow (a) *In vitro* growth phase to generate VSG AnTat1.1 negative cells. 10ml of a logarithmic uninduced population of GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* or *DYR*K RNAi cells was diluted to 6 cells/ml and seeded

in 100µl volumes in a 96-well plate. Doxycycline was added to the +DOX populations following dilution and prior to distribution in the 96 well plate. After 7 days of incubation at 37^{o}C, the cell density of the arising clones was counted using a haemocytometer. Three clones which had similar cell densities, and appeared healthy when visualised with a microscope, were selected for progression through the assay. Over the course of a further 5 days, cell counts were taken daily and the cultures are progressively scaled up by the addition of HMI-9 to prevent the populations from overgrowing. Where fresh HMI-9 was added to the induced cultures, doxycycline was 'topped up' also. On day 12, there were at least 1.25x10^{8} cells **(b)** MACS depletion of VSG AnTat1.1 positive cells. Cells were stained with an α-VSG AnTat1.1 antibody at 1:20,000. 10,000 events were captured for the unsorted (left panel) and the column-bound (far right panel) populations. Approximately 400 intact cells were detected in the MACS-sorted population (middle panel). Prior to MACS separation the population was 87.7% positive for VSG AnTat1.1. Following the depletion of VSG AnTat1.1 expressing parasites, the flow through population was 43.8% VSG AnTat1.1 negative **(c)** Amplification of VSG-specific cDNA from a post-MACS RNA sample. 10ng RNA was converted to single stranded cDNA using a primer specific to the 14mer and varying concentration of SS cDNA used for the amplification of DS VSG cDNA. 1µl of a 1:10 dilution of the SS cDNA product was sufficient to amplify a product corresponding to the size of a VSG.

Both haemocytometer counts and FACS analysis of the MACS column flow-through populations from workflow trials 1 and 2 suggested that only ~400-2000 VSG AnTat1.1 negative cells were present after depletion of the VSG AnTat 1.1 expressors. Following removal of a 1ml sample for IFA analysis, RNA was extracted and Turbo™ DNase (Invitrogen) treated using the protocols in Section 2.10. To aid in recovery of the low concentrations of nucleic acids present, the RNA was co-precipitated with the nucleic acid carrier glycogen (Thermo Scientific) (Section 2.17.3). In the presence of ethanol and salt, glycogen forms a precipitate which helps to 'trap' the target RNA. Upon centrifugation, the glycogen forms a visible pellet in which the RNA is contained, thus aiding the handling of the low concentration RNA. Unlike other nucleic acid carriers such as yeast tRNA, glycogen has been demonstrated not to interfere with downstream enzymatic or sequencing processes at concentrations of up to 8µg/µl (Thermo Scientific). Following glycogen co-

precipitation of the two post-MACS RNA samples, the lowest concentration sample (10ng RNA/µl) was used to test whether the RNA extracted was of sufficient yield and quality to amplify VSG cDNA. Hence, single stranded VSG cDNA was generated from 1µl of the 10ng/µl RNA stock using a primer specific to the 14mer (Table 2.2), and RT-PCR using the SL and 14mer primers attempted with 1µl of a 1:10 dilution of the cDNA stock, 1µl of cDNA stock or 3µl of cDNA stock (Figure 4.2c) (Section 2.11). A minus reverse transcriptase (-RT) control confirmed that the DNase treatment was effective. In all three conditions, a single tight band was visualised between 1500 and 2000bp indicating that the amplification of double-stranded VSG cDNA had been successful despite the low quantities of RNA present in the initial step.

The trial VSGseq workflows therefore demonstrated that the proposed *in vitro* growth protocol generated a sufficient number of VSG switched cells from which post-MACS RNA could be extracted and double-stranded VSG cDNA successfully amplified. Further, FACS analysis of cell populations pre- and post-MACS separation found that the MACS-sorted populations were 43.8% negative for VSG AnTat1.1, an enrichment which would aid in the detection of rare VSG variants within the population.

4.2.2 Generation and validation of pre- and post-MACS samples for VSGseq

To quantitate the expressed VSG diversity in populations of pleomorphic and inducibly monomorphic trypanosomes, the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi cell lines were tested in triplicate using the validated VSGseq workflow. For each RNAi line, the differentiation competent and differentiation defective populations were grown in triplicate at the same time, but for reasons of practicality the MACS VSG AnTat1.1 depletion step was performed separately as two groups of three on the same day.

Each growth phase lasted exactly 12 days, with the cultures being scaled-up by the addition of drug-free HMI-9 on days 7, 9 and 11 (Figure 4.3a). The mean number of population doublings during the uninduced *HYP2* RNAi cells' growth was

significantly less (p=0.0004, student's t-test) than that of the induced *HYP2* RNAi cells' (26.7 ± 0.2 vs 28.9 ± 0.3 generations). No significant difference (p>0.05, student's t-test) was found in the number of population doublings between the uninduced and induced *NEK* RNAi (25.8 ± 2.1 vs 29.2 ± 0.4 generations) or *DYRK* RNAi (28.9 ± 0.05 vs 28.7 ± 0.02 generations) populations, respectively. The number of population doublings is an estimate since the exact time when each cell count was performed was not noted; the assumption was made that 24 hours passed between each timepoint.

The knock-down of each stumpy formation quorum-sensing component was verified by qRT-PCR using Turbo™ DNase treated pre-MACS RNA (Figure 4.3b). Single-stranded cDNA was generated from 300ng total RNA using an oligo$(dT)_{20}$ primer. A gene-specific stretch of ~100bp was then amplified from 2.5µl of a 1:20 dilution of the SS cDNA using gene-specific primers (Table 2.2). Upon the induction of RNAi, the *HYP2* transcripts dropped to 0.24 ± 0.005, 0.12 ± 0.002 and 0.23 ± 0.006 relative to the uninduced levels, respectively, and the *NEK* transcripts to 0.27 ± 0.001, 0.34 ± 0.006 and 0.12 ± 0.01 relative to the uninduced levels, respectively (Figure 4.3b, left and middle panels). The transcript knock-down was not as pronounced in the induced *DYRK* RNAi line, with the transcript levels dropping to 0.66 ± 0.02, 0.33 ± 0.02 and 0.66 ± 0.05 relative to the pleomorphic uninduced populations, respectively (Figure 4.3b, right panel).

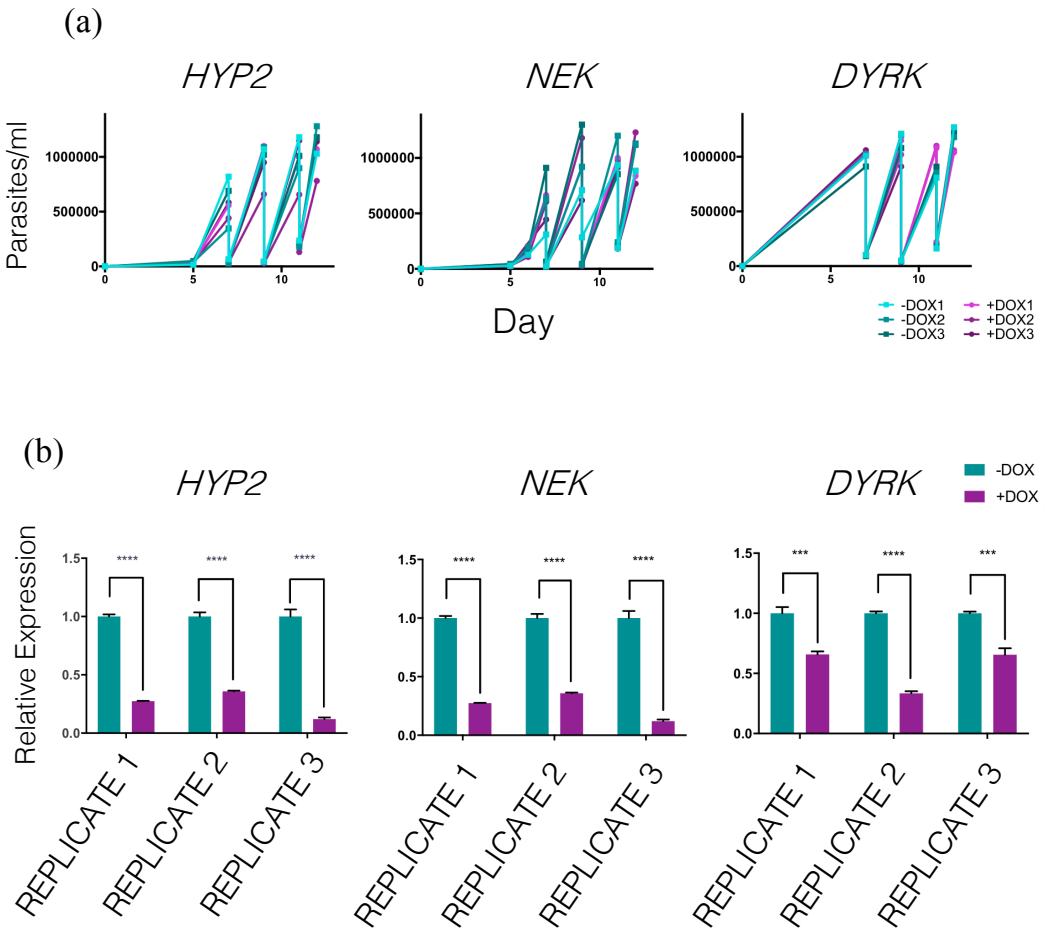


Figure 4.3 Validation of *HYP2*, *NEK* and *DYRK* knockdown in the induced GFP^{ESpro}ANTat1.1^{ES} populations (a) *In vitro* growth curves of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* (left), *NEK* (middle) and *DYRK* (right) RNAi lines (± induction) prior to pre-MACS RNA extraction, MACS depletion of VSG AnTat1.1 positive cells and extraction of post-MACS RNA. Uninduced populations are shaded in teal and induced in plum. Assays were performed once in triplicate (n=3). **(b)** Validation of *HYP2* (left panel), *NEK* (middle panel) and *DYRK* (left panel) transcript knock down upon induction. Uninduced replicates are shaded in teal, induced in plum. cDNA was generated from 300ng pre-MACS RNA. Data represents the mean ± S.D (n=3). *** = $p \leq 0.001$ **** = $p \leq 0.0001$.

Prior to labelling the populations with magnetic beads coupled to the α-VSG AnTat1.1 antibody, $1x10^6$ cells were removed, washed and fixed for FACS analysis to validate that the population originated from a VSG AnTat1.1/GFP positive cell

(Figure 4.4a) and a further $1x10^6$ fixed for cell cycle and PAD expression analysis by IFA (Figure 4.4b). The FACS profiles in Figure 4.4a show that the majority of the triplicate uninduced and induced populations for all three $GFP^{ESpro}AnTat1.1^{ES}$ RNAi cell lines were both VSG AnTat1.1 and GFP positive, strongly suggesting that the single cell from which the populations were expanded from was also doubly positive. To perform a valid comparison of the expressed VSG diversity between differentiation competent and incompetent populations, it was important that both populations remained proliferative. This was because a high proportion of stumpy cells within the populations, which are irreversibly arrested in their cell cycle, would affect the comparison since they could not form new VSG switch variants (Amiguet-Vercher et al., 2004). Proliferative cells with a 2K1N or 2K2N configuration represented ~20-25% of the populations (Figure 4.4b), and for each of the uninduced and induced RNAi lines there was no significant difference ($p>0.05$, student's t-test) in the proportion that were replicative. PAD1 protein expression was determined from the same IFA samples which were used to determine cell cycle status. In each of the $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2*, *NEK* and *DYRK* RNAi lines, the percentage of the population which expressed PAD1 was always higher in the uninduced population compared to the induced population (6.9% ± 0.8 vs 0.5% ± 0.5, 2% ± 1.8 vs 0.7% ± 1.2 and 8% ± 3.6 vs 2.7% ± 1.7, respectively), though this was only significant in the case of *HYP2* ($p=0.0003$, student's t-test). This was still only a very small proportion of the respective populations however, and since cell cycle analysis showed no accumulation in G_1, the populations were considered proliferative.

Following the MACS depletion step, the number of parasites in a 10µl sample from the column flow-through was counted on a haemocytometer. No more than 4 parasites were ever visualised in a given sample, and in 8 of the 18 flow-through samples no parasites were counted at all. A 1ml sample from the column flow through was fixed for IFA analysis of VSG expression. The number of cells in each IFA sample was variable but consistently low (no more than 17 cells were counted for a single sample) (Figure 4.4c). Prior FACS analysis had suggested that the MACS column flow-through was >40% negative for VSG AnTat1.1 (Figure 4.2b). This level of VSG AnTat1.1 depletion was attained for all of the $GFP^{ESpro}AnTat1.1^{ES}$ RNAi populations, except for the uninduced

GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi flow-through population (33.5% ± 19.1 VSG AnTat1.1 negative parasites) and the induced GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi flow-through population (14.3% VSG AnTat1.1 negative parasites) where very high levels of contamination with VSG AnTat1.1 expressing variants were observed. Since VSGseq is capable of reliably detecting variants which are expressed by 0.01% of the population however, these levels of VSG AnTat1.1 depletion were considered sufficient to improve detection of rare variants compared to the unsorted pre-MACS samples.

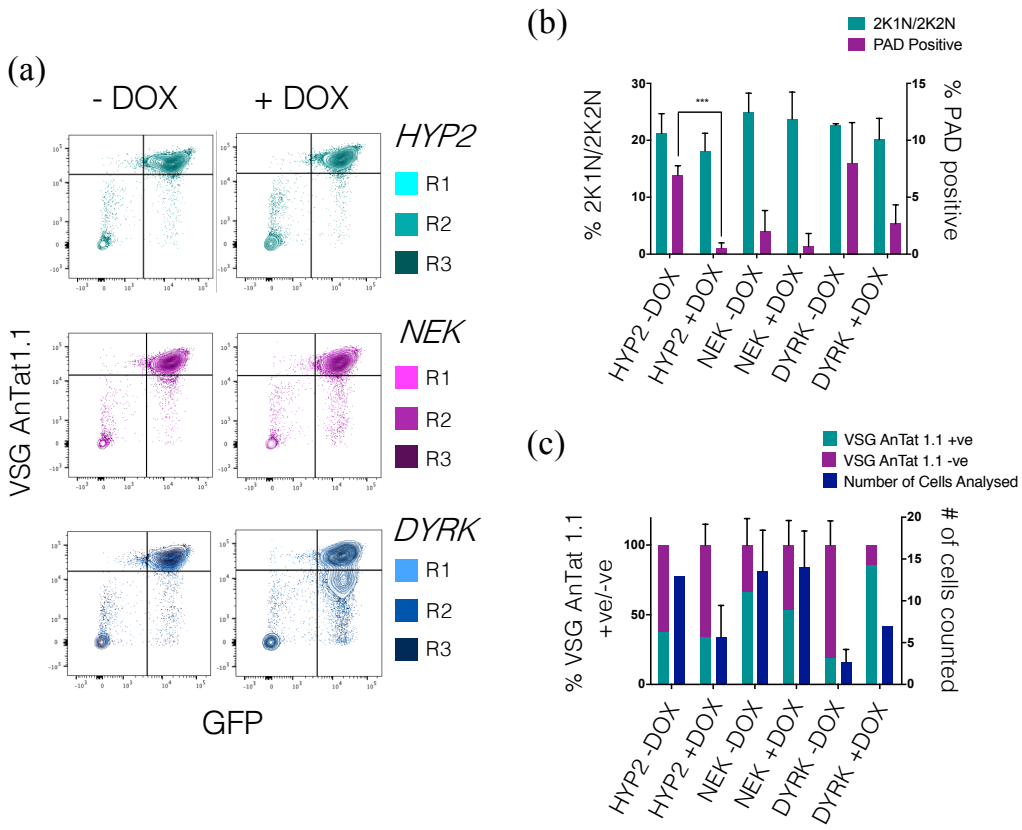


Figure 4.4 The GFP^{ESpro}AnTat1.1^{ES} populations originated from a VSG AnTat1.1/GFP positive cell and were proliferative (a) FACS analysis of VSG and GFP expression in cells fixed on day 12 prior to MACS depletion. For each replicate (R) 10,000 events were captured. VSG AnTat1.1 positivity was measured by α-VSG AnTat1.1 antibody binding and GFP positivity by fluorescence emission in the FITC channel alone. The majority of the populations were GFP and VSG AnTat1.1 positive, suggesting the single cell from which the

populations were grown from was also GFP and VSG AnTat1.1 positive. Variants which had arisen from *in situ* switches (GFP/VSG AnTat1.1 -ve) or DNA recombination (GFP +ve/VSG AnTat1.1 -ve) accumulated during the growth phase **(b)** Cell cycle and PAD1 expression analysis of paraformaldehyde fixed cells. Cells were stained with DAPI and an α-PAD1 antibody. The uninduced, differentiation competent cells did not accumulate as cell cycle arrested, PAD1 expressing stumpy forms. Data represents the mean±S.D (n=3), 250 cells were counted for each replicate. *** = $p \leq 0.001$ **(c)** IFA analysis of VSG expression in cells collected from the MACS column flow through. 1ml of sample was removed from the column flow through immediately post-MACS, washed in PBS and fixed in paraformaldehyde. Cells were stained with an α-VSG AnTat1.1 antibody and the entire sample loaded to a single well. Flow through populations were enriched for non-VSG AnTat1.1 expressors. Data represents the mean±S.D (n=3, except for the *HYP2* -DOX and *DYRK* +DOX samples where n=1 and the *NEK* -DOX sample where n=2).

4.2.3 RNA extraction and library preparation for Illumina sequencing

Pre- and post-MACS RNA was extracted, DNase treated and glycogen co-precipitated (post-MACS only) as previously described (Section 2.17.3). The RNA concentrations of the 36 samples were estimated using a NanoDrop spectrophotometer (Thermo Fisher Scientific) and ranged from 188.5-459.8ng/µl for the pre-MACS samples and 8-98.5ng/µl for the post-MACS samples (the post-MACS samples with the highest RNA concentrations corresponded to the replicates in which numerous contaminating VSG AnTat1.1 expressors were observed). The VSGseq protocol was performed in collaboration with Monica Mugnier at the Johns Hopkins Bloomberg School of Public Health (Baltimore, USA). RNA samples were shipped on dry ice and stored at -80ºC until their time of use.

First strand cDNA synthesis and the subsequent PCR amplification of VSGs was performed as described in Section 2.17.4. Following the latter, 5µl of the PCR product from 6 samples was run on an agarose gel to confirm the steps had successfully generated double-stranded VSG cDNA (Figure 4.5). A single band corresponding to the size of a VSG amplicon was visualised in all of the samples, confirming that the RT-PCR had been successful.

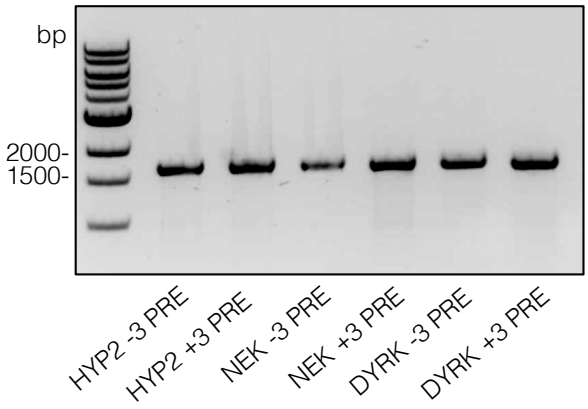


Figure 4.5 VSG Specific RT-PCR products VSG-specific RT-PCR produced a ~1600bp double stranded cDNA product, corresponding to the size of a VSG. 'Pre' refers to pre-MACS RNA, -/+ refers to the absence or presence of doxycycline.

VSG amplicons were simultaneously fragmented and tagged with universal overhangs (tagmentation) using the NextEra XT DNA Library Prep Kit (Illumina) (see Section 2.17.4). It is to these universal overhangs that specific pairs of index sequences were annealed during a further PCR amplification. As a final quality control step, small aliquots from 11 different library preparations (selected to give a broad representation of the sample set) were ran on a bioanalyzer in order to test the effectiveness of the fragmentation step (Appendix B). Typical libraries consist of a range of fragments between 250-1000bp. Each of the library preparations in Appendix B did consist of a broad range of fragments between ~200 and 1000bp, however a sharp peak at ~1000bp followed by a second large peak at ~1500bp was observed in samples 8, 9 and 10 (*DYRK* pre-MACS -DOX2/3 and *DYRK* pre-MACS +DOX1, respectively). Since the mean fragment size in these library preparations was still within the expected range however (738bp, 937bp and 697bp, respectively), the libraries were considered satisfactory. The peaks at 35 and 10380bp represent low and high molecular weight markers.

The 36 libraries were pooled such that 0.02pmol of each sample was present in a final mixture of 103.05µl. The pooled libraries were loaded onto a single Illumina

flow cell and 100bp single-end sequencing was performed in-house on an Illumina HiSeq 2500. The VSGseq data was analysed using the pipeline available on https://github.com/mugnierlab/VSGSeqPipeline, with the identified ORFs being compared by BLAST to the EATRO1125_vsgs reference database in order to identify VSG sequences (for more information see Section 2.17.5). The output MULTo analysed csv. file shows the expression of each VSG in each sample, both in terms of RPKM (reads per kilobase of transcript, per million mapped reads) and percentage of the population (RPKM for that VSG/total RPKM), and the BLAST similarity results for each given VSG compared to the EATRO 1125 reference database.

4.3 Overall VSG expression diversity is similar between pleomorphic and inducibly monomorphic populations

FACS analysis demonstrated that the majority of the parasites contained in the pre-MACS samples had not switched expression from VSG AnTat1.1 (Figure 4.4a) and this was reciprocated in the VSGseq analysis. When the uninduced and induced triplicates were combined, the mean percentage of the pre-MACS populations expressing VSG AnTat1.1 was ≥99.6% for each of the three $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines (Figure 4.6, top). IFA results, whilst limited in their usefulness due to ≤ 20 cells being analysed per replicate, had confirmed enrichment for VSG AnTat1.1 negative parasites following the MACS depletion step but also suggested that VSG AnTat1.1 expressing contaminants were present in the flow through (Figure 4.4c). The post-MACS VSGseq datasets corroborated this observation. The mean percentage of the 6 replicate post-MACS populations expressing VSG AnTat1.1 dropped to 89.8% ± 7.8, 74.4% ± 19.8 and 91.4% ± 8.8 for the $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2*, *NEK* and *DYRK* RNAi lines, respectively, representing a 34-, 85.3- and 21.5-fold enrichment for VSG switched cells.

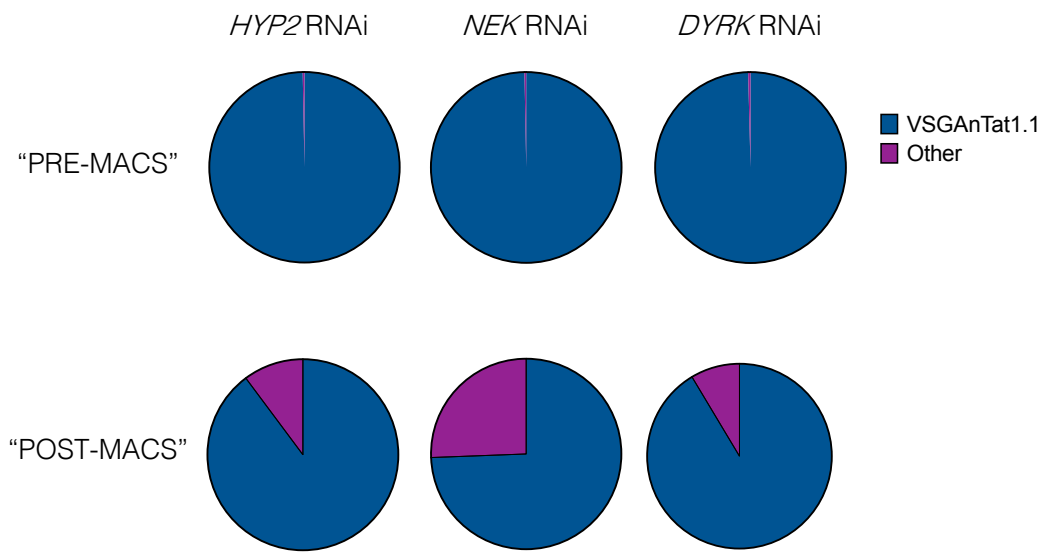


Figure 4.6 Post-MACS samples are enriched for non-VSG AnTat1.1 sequences Pie charts represent the mean percentage of reads within the combined uninduced and induced populations which mapped to VSG AnTat1.1 (blue) or any other VSG (plum), as determined by VSGseq analysis. Post-MACS *HYP2*, *NEK* and *DYRK* populations were enriched 34-, 85.3- and 21.5-fold for non-VSG AnTat1.1 expressors.

The pre-MACS samples allowed us to take an unbiased approach to examine overall VSG expression diversity in pleomorphic and inducibly monomorphic populations. As a first measure of VSG expression diversity, the numbers of different VSGs detected within the respective uninduced and induced populations were plotted alongside the percentage of the population which was made up of non-VSG AnTat1.1 expressing cells (VSG switch variants) (Figure 4.7). By accounting for both the number of VSGs and the proportion of the population which consisted of VSG switched cells, we would be able to discriminate whether induction of monomorphism had an effect on expressed VSG diversity, and therefore antigen switching. Since VSGseq does not reliably detect variants which are expressed by less than 0.01% of the population, any VSG whose expression fell below this threshold was excluded from the analysis.

The number of VSGs accounting for >0.01% of the population ranged from 3 (*HYP2* RNAi +DOX 2) up to 14 (*DYRK* RNAi +DOX 2). Upon induction of RNAi against *HYP2*, the mean number of expressed VSGs detected within the triplicate populations dropped from 7.67 ± 1.5 to 4.67 ± 1.5, whilst induction of RNAi against *NEK* or *DYRK* contrarily increased the number of detectable VSGs within the populations (7 ± 2 vs 6.7 ± 1.5 and 8.3 ± 5.1 vs 6.3 ± 1.5). None of these changes, however, were statistically significant ($p>0.05$, student's t-test). A similar number of detectable VSGs, but differences in the proportion of the total population made up of VSG switched cells, could indicate differences in either the kinetics of VSG switching, the stability of the newly activated antigen variant(s) or the growth rates of switched clones. As previously demonstrated in Figure 4.6, where the values for the uninduced and induced populations were combined, the percentage of the populations expressing VSGs other than VSG AnTat1.1 was consistently low. When considered separately, there was no statistically significant ($p>0.05$, student's t-test) difference between the percentage of VSG switched cells in differentiation competent and differentiation defective populations for any of the $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines (Figure 4.7). Together, these results suggest that the induction of monomorphism does not reduce expressed VSG diversity compared to the parental pleomorphic population. Since the number of expressed VSGs, and the proportion of the population consisting of switched cells, remains consistent between the two groups, this suggests that the pleomorphs and inducible monomorphs are generating new VSG variants at a similar rate.

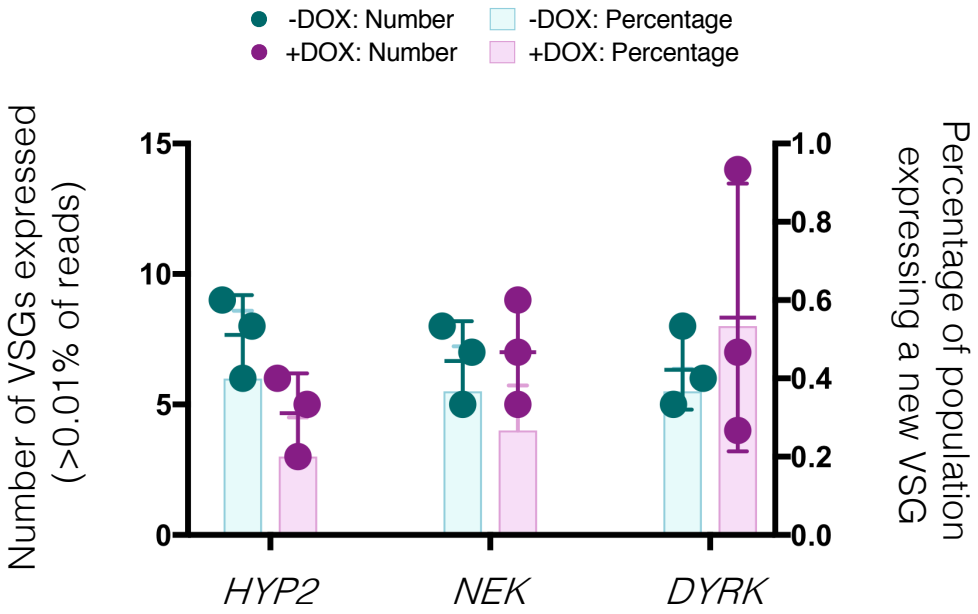


Figure 4.7 Induction of monomorphism does not reduce VSG expression diversity The number of VSGs (dot symbols) expressed by over 0.01% of the total population (reliability threshold) was counted for the uninduced (-DOX, teal) and induced (+DOX, plum) GFP^{ESpro}AnTat1.1^{ES} *HYP2* (left), *NEK* (middle) and *DYRK* (right) RNAi lines. The percentage of cells in these populations that consisted of non-VSG AnTat1.1 expressors is represented by the coloured bars. Induction of monomorphism did not significantly reduce the number of detectable VSGs, nor the prevalence of VSG switched cells within the population, suggesting that the uninduced and induced populations are switching their expressed VSGs at a similar rate. Data represents the mean±S.D. (n=3).

Although the overall numbers of expressed VSGs within the pleomorphic and inducibly monomorphic populations were similar, this did not rule out the possibility that induction of monomorphism was influencing which VSGs were being selected for expression. Monomorphic laboratory-adapted cells are suggested to preferentially change expressed VSGs via *in situ* switches whilst pleomorphic populations more often employ DNA recombination (Liu et al., 1985; McCulloch et al., 1997; Robinson et al., 1999). The repertoires of expressed VSGs within the pre-MACS datasets were explored for VSGs unique to either uninduced or induced conditions to investigate if induction of monomorphism changed VSG expression bias (Figure 4.8). Overall, 20 different VSGs were present at a frequency of >0.01% in the 6

populations, half of which were present in at least one replicate of both the uninduced and induced populations (Figure 4.8, 'combined'). No single VSG transcript appeared in all three replicates of all three GFP^{ESpro}AnTat1.1^{ES} RNAi lines. Three VSGs were unique to at least one of the uninduced, pleomorphic replicates, whilst more than twice this were unique to at least one of the inducibly monomorphic replicates. Although this could seem to suggest that the induction of monomorphism can drive the monomorphic populations to select distinct VSGs to those in the pleomorphic populations, these results are biased by the +DOX2 *DYRK* RNAi replicate in which 2-3x more VSG types were identified than in any other replicate (Figures 4.7 and 4.8, *DYRK* RNAi). Coincidently this replicate was the only *DYRK* RNAi replicate which saw levels of transcript knock down over 50% (Figure 4.3b). On balance, it is likely that the numbers of unique VSGs in the inducibly monomorphic and pleomorphic populations were more similar overall, as seen in the *HYP2* and *NEK* RNAi lines (Figure 4.8, *HYP2* and *NEK* RNAi). Of the VSGs 'unique' to either the uninduced or induced conditions, only two (DN171_c0_g1_i1_1_22_1453 and DN480_c0_g1_i3_1_45_1416) were expressed in more than one replicate (*HYP2* -DOX 1/3 and *HYP2* -DOX 1/2/3, respectively). Since both these VSGs were present in the pleomorphic and inducibly monomorphic *NEK* and *DYRK* RNAi populations, they do not seem to be preferentially activated by pleomorphic populations.

The induction of monomorphism was not associated with the selection of VSGs distinct from the those activated in the pleomorphic populations. Therefore, I next asked if the proportions of certain VSGs were always higher in one population compared to another. For this, I studied those VSGs which were consistently expressed in every triplicate within a given condition, rather than those which appeared transiently in a single replicate. Figure 4.8b shows that no one VSG predominated in every triplicate demonstrating, even in these pre-enrichment sample sets, that a diverse range of VSGs are being expressed by both the different populations, and within the triplicates themselves. Indeed, no consistent pattern of expression was observed within any of the uninduced or induced populations of the three GFP^{ESpro}AnTat1.1^{ES} RNAi lines. Apparently, therefore, any differences in VSG expression between the inducible monomorphs and pleomorphs was a product of

their inherent diversity, and not caused by the induction of RNAi against a gene involved in stumpy formation.

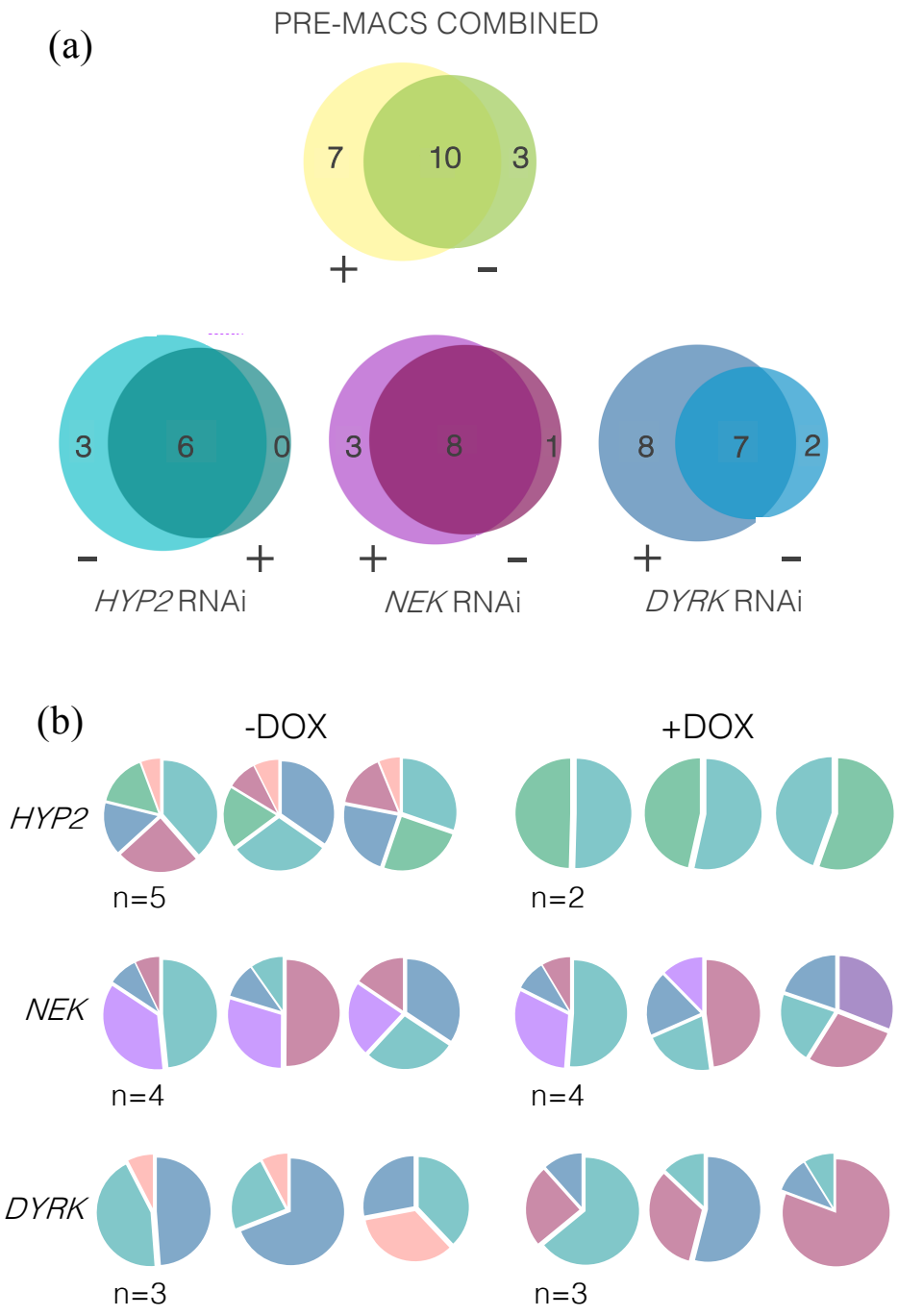


Figure 4.8 The majority of the VSGs identified in the pre-MACS datasets were expressed by both the pleomorphs and inducible monomorphs (a) Venn diagrams showing intersections of expressed VSGs in the pre-MACS datasets. The 'combined' dataset incorporates the data from the *HYP2*, *NEK* and *DYRK* GFP^{ESpro}AnTat1.1^{ES} lines. -/+ refers to

the addition of doxycycline **(b)** Pie charts showing the proportion of reads which map to VSGs occurring in every triplicate for a given condition (- or + DOX) in the pre-MACS datasets. VSG AnTat1.1 was excluded from the analysis. Each VSG is represented by a different colour, the colours are consistent between each condition. The number of VSGs is denoted by 'n'.

4.4 Post-MACS analysis shows that the induction of monomorphism is not associated with the selection of specific VSG genes

VSGseq analysis of the pre-MACS samples demonstrated that the induction of monomorphism does not reduce the diversity of VSGs expressed within the population (measured by the number of VSGs identified within the populations), change the identity of the VSGs selected for expression or influence the extent to which certain VSGs are represented within the population compared to the pleomorphic populations. However, no VSG other than VSG AnTat1.1 was reported to be expressed in >0.26% of any of the populations and, therefore, many variants expressed at very low levels within the population could have fallen below the pre-MACS sequencing threshold. Therefore, MACS depletion of VSG AnTat1.1 expressing parasites was carried out facilitating the identification of rare variants within the populations. As highlighted earlier, MACS depletion enriched for AnTat1.1 expressors in the GFP^{ESpro} $AnTat1.1^{ES}$ *HYP2*, *NEK* and *DYRK* RNAi populations (uninduced and induced combined) 34-, 85.3- and 21.5-fold, respectively (Figure 4.6).

The application of MACS to enrich for VSG switched parasites complicates the use of VSG counts as a method of comparison between inducibly monomorphic and pleomorphic populations, as was performed in Section 4.3. For example, a higher number of VSGs identified within the populations could indicate either a higher level of expressed VSG diversity, or simply that the MACS depletion step had been more efficient for a given replicate. However, since there was no significant difference ($p>0.05$, student's t-test) in the mean percentage of reads accounted for by

VSG switched cells between uninduced and induced *HYP2*, *NEK* and *DYRK* RNAi lines (6.97% ± 5.1 vs 13.5% ± 9.7, 24.04% ± 15.6 vs 27.1% ± 26.9 and 6.04% ± 5.3 vs 11.2% ± 12.1 VSG AnTat1.1 negative reads >0.01% population), it was assumed that the MACS depletion step had worked similarly as well between the uninduced and induced populations of each RNAi line.

Overall, 106 individual VSGs were identified in the post-MACS datasets, representing a 5.3-fold increase on the number identified by the pre-MACS analysis. The percentage of each replicate population (VSG RPKM/total population RPKM) expressing a given VSG was $\log_{10}$ transformed and the results expressed in the heatmaps in Figure 4.9a. VSG AnTat1.1 appears as a dark red band at the top of all the heat plots, indicating that it is still the most prevalent VSG sequenced in the populations despite the MACS depletion step. Before considering the differences in VSG gene expression between the uninduced and induced populations, it was observed that even within the uninduced or induced triplicates themselves there were high levels of VSG expression diversity. For each condition, the percentage of VSGs which appeared in all three triplicates is shown in the horizontal bar chart below the heatmaps, and ranges from only 13% (*DYRK* RNAi inducible monomorphs) to 39% (*NEK* RNAi pleomorphs).

The identity of these consistently expressed VSGs, and the proportion to which they appeared in their respective populations, was investigated in order to determine a) whether any were more common to pleomorphic or inducibly monomorphic populations, or b) whether they were expressed to a greater extent in one population compared to another (Figure 4.9b). VSG AnTat1.1 reads were again excluded from the analysis to increase resolution. As was observed in the pre-MACS analysis, there was no consistent pattern of VSG expression either within the triplicates, or within the pleomorphic and inducibly monomorphic populations themselves. Only 2 assembled VSGs were present in every uninduced and induced population in all three of the $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines. The first of these, DN921_c0_g1_i1_1_1_1543, never surpassed expression by >0.22% of any of the populations however the second, DN30_c0_g1_i2_1_2_1484, was expressed by a third of one of the induced $GFP^{ESpro}AnTat1.1^{ES}$ *NEK* RNAi total populations (note

that these were calculated when VSG AnTat1.1 was included in the analysis; the relative proportions are exaggerated in the pie charts in Figure 4.9b since VSG AnTat1.1 was removed for these). DN103_c0_g1_i1_1_16_1477 was the only example in the post-MACS datasets of a VSG which was the most highly expressed VSG after VSG AnTat1.1 in every replicate within a condition (*DYRK* RNAi +DOX1/2/3) (Figure 4.9a, purple box and Figure 4.9b).
DN103_c0_g1_i1_1_16_1477 was also the most highly expressed VSG after VSG AnTat1.1 in one other inducibly monomorphic replicate (*HYP2* RNAi +DOX3), however it otherwise appeared at various percentages within both pleomorphic and inducibly monomorphic populations, suggesting that it is not preferentially activated following the induction of monomorphism. There was only one instance where a VSG was consistently expressed in all three uninduced replicate populations at levels 10-fold greater than all three induced replicates (mean 1.42% ± 0.47 in the uninduced populations and 0.07% ± 0.05 in the induced populations) (Figure 4.9a, black box and Figure 4.9b). DN104_c0_g1_i2_1_0_1523 was not a direct match to any of the VSGs contained on the EATRO 1125 database, but showed 84% identity over 86% of its length to Tb1125VSG-276. No VSGs were consistently expressed at levels ≥10-fold greater in the inducibly monomorphic populations compared to the pleomorphic populations for any of the three RNAi cell lines and furthermore, no VSG was ever found in all three replicates of either the uninduced or induced population and not at all in the corresponding uninduced or induced populations. These observations therefore corroborate the pre-MACS analysis in Section 4.3, and demonstrate that the induction of monomorphism is not associated with a change in VSG expression bias.

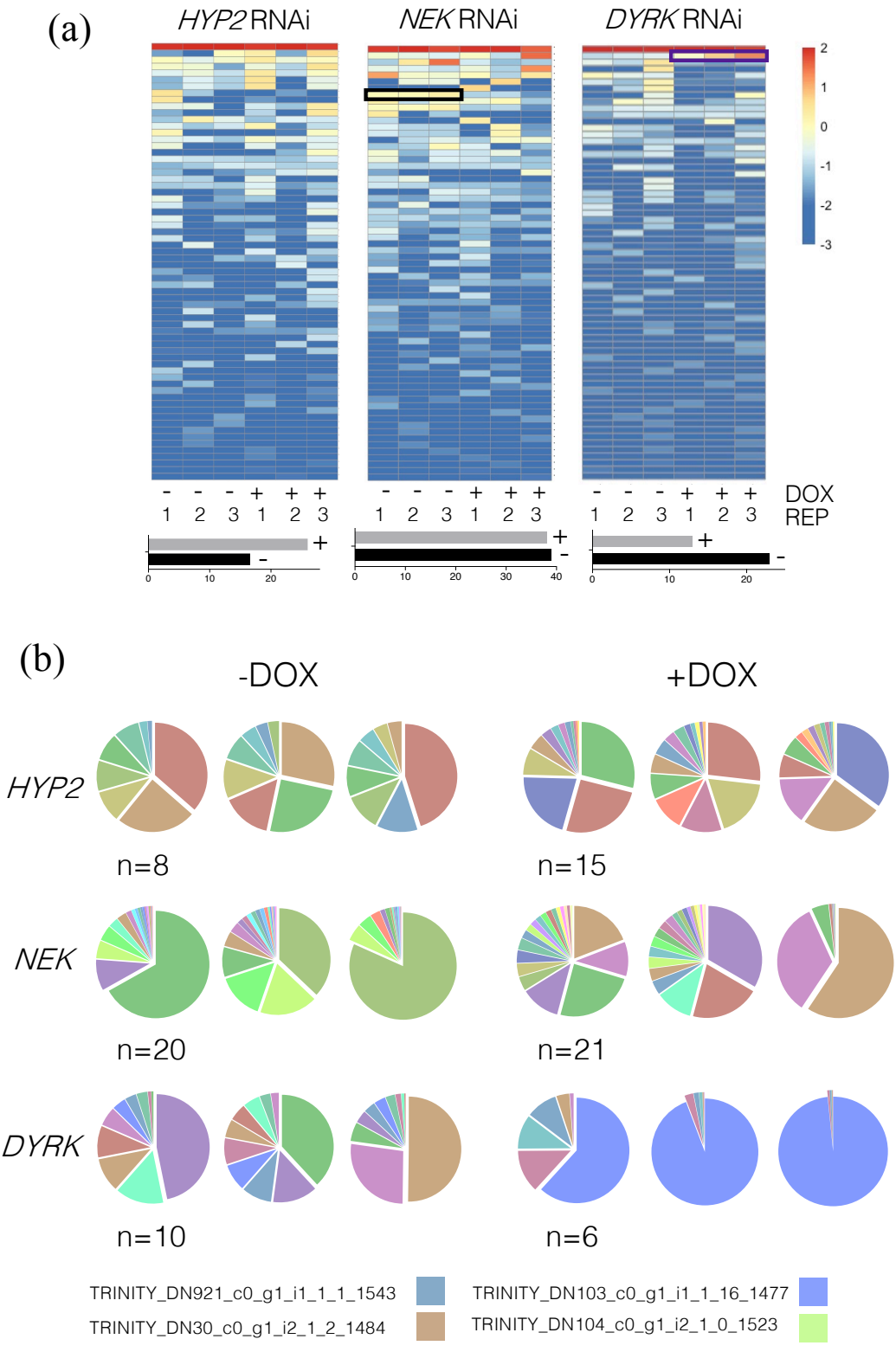


Figure 4.9 VSG expression in the post-MACS samples is highly diverse (a) VSG expression heatmap showing the percentage (log_{10} transformed) of each replicate of the

uninduced and induced *HYP2*, *NEK* and *DYRK* RNAi populations expressing a given VSG. The percentage of the population expressing a given VSG was calculated by the VSG specific RPKM divided by the total population RPKM. Where a VSG was not expressed in a given replicate (i.e. expression = 0), 0.001 was added in order to allow for the $\log_{10}$ transformation. Each row represents a different VSG. The black box around three *NEK* RNAi -DOX replicates indicates the only instance where VSG expression was 10-fold greater in all three population replicates for one condition (for e.g. -DOX) than the other (for e.g. +DOX). The purple box around three *DYRK* RNAi +DOX replicates indicates the only instance where a VSG was consistently the highest expressed VSG (excluding VSG AnTat1.1) in all triplicates for a given condition (-DOX or +DOX). The colour intensity scale on the right ranges from dark red (100% of the population) to dark blue (0.001% of the population). The order of the listed VSGs is consistent between the uninduced and induced samples, but not between the three GFP^{ESpro}AnTat1.1 lines. For each GFP^{ESpro}AnTat1.1^{ES} RNAi line, the number of VSGs which were identified in all the uninduced (black) or induced (grey) triplicates is shown as a percentage of the total VSG population in the bar charts below each heat plot. "–" and "+" represents the absence or presence of doxycycline. **(b)** Pie charts showing the number of reads which map to VSGs occurring in every triplicate for a given condition in the post-MACS datasets. VSG AnTat1.1 was excluded from the analysis. Each VSG is represented by a different colour (the colours are consistent between each condition and the pie-charts in Figure 4.8). The number of VSGs is denoted by 'n'. The VSGs mentioned in the text are shown in the key at the bottom.

Mugnier et al. (2015) used VSGseq analysis to quantitively analyse VSG expression during *in vivo* infections of mice, and demonstrated that high levels of VSG expression diversity were present throughout the populations, even within the parasitaemic valleys. Of the 192 variants that were identified during their study, the majority of these (86%) appeared in more than one infection and almost half (46%) appeared in all. Of the VSGs identified in the post-MACS datasets, 60 (56.6%) appeared in at least one of both the uninduced and induced GFP^{ESpro}AnTat1.1^{ES} RNAi populations (Figure 4.10, combined) and only 2 (1.9%) were assembled in every replicate of every condition in each of the GFP^{ESpro}AnTat1.1^{ES} RNAi populations. Higher levels of VSG expression diversity in an *in vitro* setting is likely explained by a lack of immune selection against variants. Alternatively, the greater levels of diversity seen in this study may be because Mugnier et al. (2015) did not

deploy MACS to rid their populations of the initiating VSGs, or because of the inherent VSG instability of the AnTat1.1 90:13 cell lines used (Chapter 3 has already demonstrated that the cell lines switch their expressed VSG at rates of ~$2x10^3$-$5x10^4$ switches/cell/generation). When studying the intersections of expressed VSGs in pleomorphs and inducible monomorphs for each of the three $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines individually, the majority of the VSG genes expressed by the *HYP2* and *NEK* RNAi lines were found in both populations (59% and 58%, respectively) (Figure 4.10, *HYP2*, *NEK*). However, as was observed for the pre-MACS datasets, less than half of the VSG genes (38%) expressed by the *DYRK* RNAi line were common to both the pleomorphic and inducibly monomorphic populations (Figure 4.10, *DYRK*). Unlike our observation in Figures 4.7 and 4.8a, in this case the low percentage of shared VSGs between the pleomorphs and inducible monomorphs could not be explained by a single replicate in one condition expressing 2-3X more VSGs than any other replicate. None of the VSGs unique to either the *DYRK* RNAi pleomorphs or inducible monomorphs were present in triplicate, and only one VSG was present at >1% in the population (DN112_c0_g1_i3_1_4_1501 was expressed by 2.7% of a pleomorphic *DYRK* RNAi population). Since these VSGs are expressed at such low levels within the populations, and 87.5% are expressed in only one replicate, the differences in VSG expression observed in Figure 4.10 between the *DYRK* RNAi pleomorphs and inducible monomorphs were likely the result of the inherent antigen instability of this line of trypanosomes and the semi-random nature of VSG switching, not the loss of differentiation capacity. The VSGs unique to either the inducibly monomorphic or pleomorphic populations for each of the $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2*, *NEK* or *DYRK* RNAi lines were distributed approximately evenly between the two populations. This suggests, that while there may have been some differences in which VSGs were selected for expression, the new variants were being generated at the same rate in the pleomorphic and inducibly monomorphic populations.

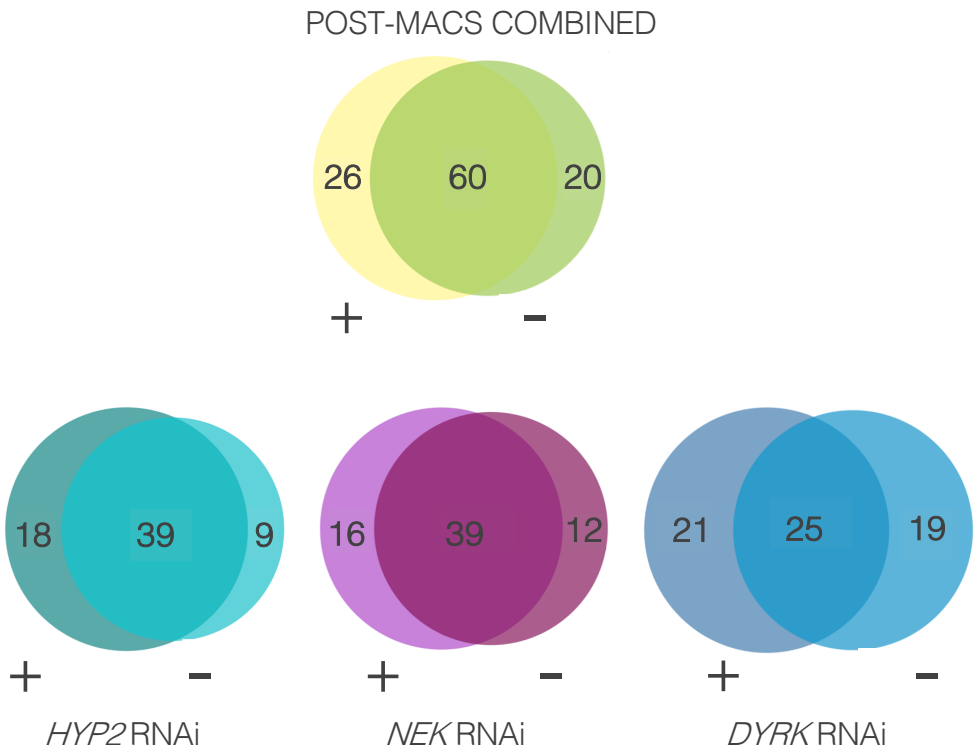


Figure 4.10 *In vitro* populations of *DYRK* RNAi pleomorphs and inducible monomorphs are highly diverse in their VSG expression but appear to generate new variants at roughly the same rate Venn diagrams showing intersections of expressed VSGs in the post-MACS dataset. The 'combined' dataset incorporates the data from the *HYP2*, *NEK* and *DYRK* GFP^{ESpro}AnTat1.1^{ES} lines. -/+ refers to the addition of doxycycline.

Overall, pre- and post-MACS VSGseq analysis of the triplicate pleomorphic and inducibly monomorphic GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi populations identified 107 different VSGs (Figure 4.11a). Just under 60% of the VSGs were shared between at least one replicate in a pleomorphic and inducibly monomorphic population. The remainder of the VSGs were evenly distributed between the two groups- 18% were only assembled in a pleomorphic population and the remaining 23% were only assembled in an inducibly monomorphic population. Diverse VSG expression was witnessed at all stages of the analysis, both within the triplicates of a given uninduced or induced RNAi population, and between the inducibly monomorphic and pleomorphic populations themselves. In addition to this,

and perhaps somewhat surprisingly, VSG expression was not particularly consistent between the three GFP^{ESpro}AnTat1.1^{ES} RNAi lines either, despite these having the same AnTat1.1 90:13 genetic background. Of the 107 assembled VSG genes, less than half (40.1%) were expressed by all three of the GFP^{ESpro}AnTat1.1^{ES} RNAi lines. The GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi cell line appeared to exhibit the most diverse VSG gene expression; 12% of the VSGs expressed within its populations were never expressed in a *NEK* or *HYP2* RNAi line for the duration of the experiment (Figure 4.11b). This could signify that the *HYP2* RNAi line has started to become divergent in its VSG expression compared to the *NEK* and *DYRK* RNAi lines during their different laboratory passage, or that during the growth phase of the experiment some populations underwent an extra VSG switch. The pleomorphic and inducibly monomorphic *HYP2* RNAi populations did not undergo a greater number of population doublings compared to the *NEK* or *DYRK* RNAi populations, though FACS-based VSG switch assays did demonstrate that switching occurred more frequently in this line compared to the *NEK* RNAi line (~2x10^{-3} vs ~5x10^{-4} switches/cell/generation) (Chapter 3). No VSGs unique to either the *NEK* or *DYRK* RNAi lines were assembled.

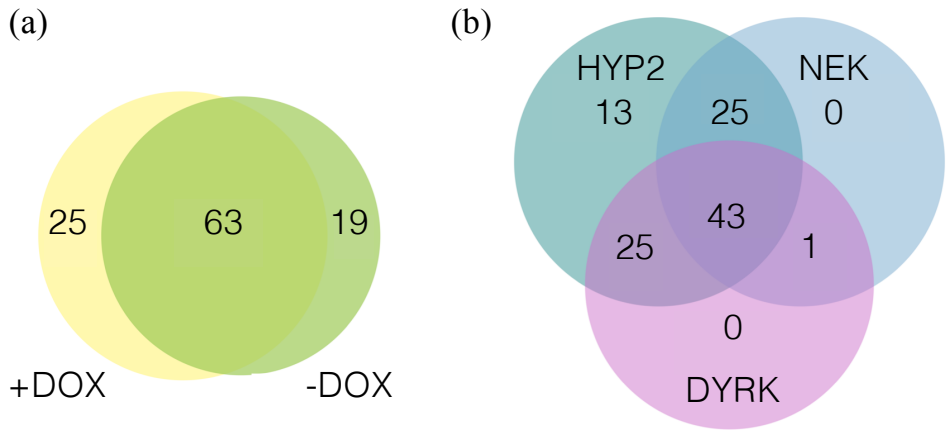


FIGURE 4.11 The GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi line expresses VSGs not assembled in the *NEK* and *DYRK* RNAi lines Venn diagrams showing intersections of expressed VSGs **(a)** in the inducibly monomorphic and pleomorphic populations (pre-and post-MACS data for each of the three RNAi lines is combined) and **(b)** in the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines (pre- and post-MACS data for each of the pleomorphic and inducibly monomorphic populations is combined). -/+ refers to the addition of doxycycline.

4.5 *In vitro* populations of trypanosomes do not readily express mosaic VSGs

Mosaic VSGs, formed by segmental gene conversion, are critical to the expansion of a parasite's VSG repertoire and therefore facilitate the maintenance of long-term infection (Mugnier et al., 2015). Although it was not anticipated that in an *in vitro* setting, and after only 12 days passage, mosaic sequences would be identified at a detectable level, the datasets were nevertheless searched for the presence of potential mosaic VSGs. Within their datasets, Mugnier et al. (2015) defined mosaic VSG candidates as having <80% of their length aligning to any sequence within two independently assembled EATRO 1125 genomes and as being present in only one infection at >0.1% of the population (for the purpose of this *in vitro* study, this was interpreted as one culture replicate). Three mosaics were validated by PCR as being present within the gDNA of the parasites collected from the infection, but absent in the gDNA of the parental strain (Mugnier et al., 2015). As expected, however, no assembled VSGs from the inducibly monomorphic or pleomorphic populations fitted these criteria.

4.6 Discussion

It has long been reported that laboratory-adapted, monomorphic strains of *T. brucei* switch their expressed VSG at a significantly lower rate than differentiation competent pleomorphic trypanosomes (Turner, 1997). Despite the wide-spread nature of this claim, it has never been demonstrated whether the reduction of antigen switch frequency and loss of differentiation capacity are directly linked, or whether the two processes are simply co-selected during long-term passage. Our results in Chapter 3 showed that the induction of monomorphism, mediated by the inducible knock-down of a gene involved in the stumpy formation pathway, did not significantly reduce the frequency of antigenic variation nor significantly change the proportion of switches mediated by DNA recombination compared to the respective

pleomorphic populations. This suggested that pleomorphism and antigen switch frequency can be uncoupled. In this Chapter we have explored a second method for establishing whether the induction of monomorphism directly influences the rate of antigenic variation- VSGseq. Developed by Mugnier et al. (2015), VSGseq is a targeted RNA sequencing approach whereby primers specific to the conserved 5' and 3' ends of all mature VSG mRNAs are used to amplify VSG cDNA. Illumina sequencing and *de novo* sequence assembly follows, providing a comprehensive view of which VSGs were being expressed, and at which frequency, within the population at the time of sampling.

Triplicate populations of pleomorphs and inducible monomorphs were grown from a single VSG AnTat1.1/GFP positive cell *in vitro* for each of the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines. For each of the final populations, both total population RNA (pre-MACS) and RNA enriched for non-VSG AnTat1.1 sequences following MACS depletion of VSG AnTat1.1 expressing parasites (post-MACS) was extracted and sequenced. VSGseq analysis of pre-MACS RNA samples allowed for an unbiased quantification of VSG expression diversity within the pleomorphic and inducibly monomorphic populations. For each of the RNAi lines, induction of monomorphism was shown to not reduce the number of VSGs identified, nor the extent to which these VSGs were represented within the population, compared to the pleomorphic population. To discriminate if the induction of monomorphism had influenced the selection of VSGs for expression, the post-MACS GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi datasets, which are enriched 34-, 85.3- and 21.5-fold for reads which do not map to VSG AnTat1.1, were analysed. The analysis found that there were no VSGs which were preferentially activated in pleomorphic or inducibly monomorphic populations. Consequently, any differences observed in VSG expression diversity were due to the inherent instability of VSG expression in these trypanosome lines and the semi-random nature of VSG switching, not the induction of monomorphism. Overall, these results provide further evidence that VSG switch frequency and differentiation competency can be uncoupled.

On the most basic level of comparison, the conclusions drawn from the VSGseq analysis support the evidence presented in the FACS-based VSG switch assay experiments namely that antigen switch frequency and differentiation capacity can be uncoupled. The induction of monomorphism did not significantly reduce the number of VSGs identified in the pre-MACS datasets, nor the extent to which non-VSG AnTat1.1 VSGs were represented in the population, in the same way that FACS-based VSG switch assays did not detect a significant drop in the proportion of the populations staining negatively for VSG AnTat1.1. The FACS-based VSG switch assays furthermore found that there was no significant change in the preferred method of VSG switching (i.e. transcriptional switches or DNA rearrangements) following the induction of monomorphism. In support of this, the VSGseq analysis did not detect a shift in VSG expression towards a 'monomorph-specific' VSG expression profile.

In Chapter 3, it was demonstrated that the VSG switch rate increased approximately 3.5-fold following the knock-down of *HYP2*. The evidence presented in Figure 4.7 did not corroborate this observation and, in fact, the induced GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi populations were found to contain the lowest mean number of detectable VSGs (4.67 ± 1.5 VSGs) of all the conditions and reporter cells tested. It was discussed in Section 3.10 that *HYP2* could contribute to the maintenance of monoallelic expression, such that when the RNAi was induced, cells began to express more than one VSG on their surface, diluting the presence of VSG AnTat1.1 on the surface of the cells and therefore leading to them being gated as negative in the VSG switch assays. If *HYP2* RNAi did indeed lead to a loss of monoallelic control, the transcript levels from dysregulated BESs would be so low that it would be very unlikely that they would be identified at the population level. This would explain why we did not measure a higher number of VSGs being expressed in the induced GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi populations. Since a repertoire of antibodies targeting *T. brucei* AnTat1.1 90:13 ES VSGs was not available to us, we could not test this hypothesis by IFA. As previously suggested, single cell sequencing of a number of induced *HYP2* RNAi clones could provide clarity.

As discussed in Sections 1.2.7 and 3.10, a limitation of the FACS-based VSG switch assay is that within the VSG AnTat1.1 negative staining population it is impossible to know the phenotype or genotype of the cells. Are each of the VSG AnTat1.1 negative cells the product of multiple switch events? Or alternatively are they the progeny of an early switch event to another stably expressed VSG? Furthermore, are these the result of primary or secondary VSG switches? And is the same VSG being activated by different methods? Whilst the VSGseq analysis does not answer the latter two questions (for this the VSG switched cells would have to be cloned and analysed by PCR and Southern Blot (Hovel-Miner et al., 2012)), it does address whether the VSG AnTat1.1 negative staining cells within the pleomorphic and inducibly monomorphic populations are similar with regard to their VSG expression. Since this was indeed the case, the VSGseq analysis thereby adds weight to the conclusions formed in Chapter 3.

Both the uninduced and induced GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi populations, and the uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi population, were consistently shown to switch expression of their VSGs at rates ~4-fold less than the GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi populations via FACS-based switch assays (Figure 3.14). The results in Figure 4.7 do not obviously present evidence for these lines switching their expressed VSGs less frequently than the others. In fact, compared to the induced *HYP2* RNAi cells, which were reported to have a significantly greater VSG switch rate, they consist of a greater proportion of VSG AnTat1.1 negative cells and express a greater number of VSGs. In theory, if a population switches VSG expression at a high rate, the number of different VSGs assembled will be greater than that of a population switching VSG expression at a low rate. Hovel-Miner et al. (2012) determined the expressed VSG diversity in populations of TERT$^{-/-}$ mutants for which they had already calculated the observed switching frequency (OSF) and found that whilst a small trend towards increased depth of VSG expression with increased OSF, there was not a linear relationship. This was demonstrated when sequencing identified only one expressed VSG within the population of cells with the highest OSF (Hovel-Miner et al., 2012). The VSG expression diversity in this case, however, was determined following the subcloning of VSG RT-PCR products into pGEM-T Easy (Promega) before sequencing and NCBI BLAST identification of

sequences. Subcloning of RT-PCR products, as opposed to VSGseq, could introduce bias for a highly expressed variant and therefore prevent the identification of rarer VSG switch variants within the population. That only 4 VSGs were identified in a $TERT^{-/-}$ population which switched VSG expression with an OSF of $2.1x10^{-4}$, despite enrichment for switched cells via MACS, could be a testament of the reduced depth of this method. For example, in the uninduced $GFP^{ESpro}AnTat1.1^{ES}$ *NEK* RNAi cell line which switched at a rate of $6.3x10^{-4}$ switches/cell/generation, between 35 and 37 VSGs were identified by VSGseq. Within their population of frequently switching $TERT^{-/-}$ mutants however, Hovel-Miner et al. (2012) discovered that the low diversity of VSG expression was the result of multiple independent VSG switch events generating the expression of the same VSG. Figure 4.9b, in which the identity of the consistently expressed VSGs within the populations were analysed, does show that a higher number of these common VSGs were expressed within the uninduced and induced *NEK* RNAi populations compared to the *HYP2* and *DYRK* RNAi populations. However, this result is likely due to the increased detection of rare variants within the *NEK* RNAi populations due to the more efficient MACS depletion of VSG AnTat1.1 expressing cells (Figure 4.6). Thus, overall both the *NEK* RNAi populations and the uninduced *HYP2* RNAi population expressed as many different VSGs (and to a similar extent within the population) as the *DYRK* RNAi line, which had a ~4-fold greater VSG switch rate. Further, post-MACS VSGseq analysis of the VSGs activated within the population did not provide strong evidence to support the hypothesis that the more frequently switching *DYRK* RNAi line was reactivating a given VSG more often than the slower switching *NEK or HYP2* RNAi lines. Therefore, there may be limitations to using expressed VSG diversity as a proxy for VSG switch rate. However, within the context of this study, the important comparison was whether the induction of monomorphism either reduced the numbers of VSGs identified, decreased the extent to which switched VSGs were represented within the population or changed the VSG expression hierarchy compared to the corresponding uninduced pleomorphic population. For each of these questions, the answer remains no.

In Chapter 3, GFP expression was used as a marker of the transcriptional activity of the VSG AnTat1.1 ES. It has been suggested that pleomorphic cells

preferentially switch expressed VSG using DNA rearrangements while monomorphs more often switch their transcribed ES (Liu et al., 1985; McCulloch et al., 1997; Robinson et al., 1999). In this Chapter I found that, after the *in vitro* growth phase, the majority of the uninduced and induced populations for all three GFP^{ESpro}AnTat1.1^{ES} RNAi lines remained GFP positive. This implies that the lines were mainly switching expressed VSG by DNA recombination rather than by using an *in situ* switch. In each of the post-MACS datasets, no fewer than 24 individual VSGs were found to be expressed within a single GFP^{ESpro}AnTat1.1^{ES} *HYP2* or *NEK* RNAi replicate. Since this is greater than the estimated number of sub-telomeric BESs (~15-20), the VSGseq analysis further implies that the inducible monomorphs and pleomorphs do switch their expressed VSG by DNA rearrangements. However, without knowing the identity of the ES resident VSG genes in the EATRO 1125 cell line, the extent to which DNA recombination or ES switches are happening cannot be established by this analysis alone. In three *DYRK* RNAi replicates (one uninduced and two induced), the number of VSGs assembled within the populations dropped to those which could be consistent with expression solely from ESs (19, 13 and 20). However, the lower numbers of VSGs identified within these *DYRK* RNAi populations are unlikely to signify preference for *in situ* switches of transcription since it was within these exact replicates that VSG AnTat1.1 contamination of the MACS flow-through was highest (97.4%, 98.7% and 92.4% of reads mapping to VSG AnTat1.1, respectively). Therefore, it is likely that many of the rarer variants within the populations were not detected. In the remaining two uninduced *DYRK* RNAi replicates 23 and 35 VSGs were detected, respectively, and 33 in the remaining induced population, thus confirming the presence of switches mediated by DNA recombination. Full annotation of the EATRO 1125 VSGnome and subtelomeric ESs, as is available for the Lister 427 strain (Cross et al., 2014; Hertz-Fowler et al., 2008; Muller et al., 2018), would allow us to more confidently determine whether VSG switches were mediated by *in situ* switches of transcription or DNA rearrangements. At present however, this information is not available. Sequence-based quantifications of VSG expression during chronic infections of rodents have demonstrated that VSG expression is highly diverse, even within the parasitaemic valleys (Hall et al., 2013; Mugnier et al., 2015). Compared to the 20

VSGs which were expressed in our 18 pre-enrichment cultures, Hall et al. (2013) (via VSG subcloning and sequencing) sequenced ~800 VSGs from 11 infections over a period of ~30 days, whilst Mugnier et al. (2015) (via VSGseq) assembled 192 VSG sequences from 4 infections over a period of 30 days (Hall et al., 2013; McCulloch and Field, 2015; Mugnier et al., 2015). On average, Hall et al. (2013) and Mugnier et al. (2015) (neither of whom enriched their populations for VSG switched parasites prior to the isolation of RNA) measured 20 and 28 variants per peak of parasitaemia, respectively. Whilst this is far greater than the average number of variants detected by RNAseq in our pre-MACS samples (~7), Mugnier et al. (2015) observed that the same VSGs occurred frequently within the different infections. 86% of the assembled VSGs appeared in more than one infection whilst almost half appeared in every infection. In contrast, only 1.9% of the VSGs assembled in this study (pre- and post-MACS combined) were present in every sequenced replicate. That such VSG expression diversity occurs between the different *in vitro* populations could be due to the absence of immune selection. In chronic infections, pre-existing cross-reacting antibodies can limit the effective VSG repertoire by clearing newly generated variants before they have a chance to establish within the population. *In vitro*, however, this immune selection is absent and the parasites are free to exploit their full VSG archive. This 'freedom of expression' and the inherent semi-random nature of VSG switching, could explain why different replicates of the same cell line exhibit such diversity in their VSG expression.

Hall et al. (2013) and Mugnier et al. (2015) both studied expressed VSG diversity over periods of approximately 30 days in chronic mouse models. In contrast, our study has investigated VSG expression diversity in *in vitro* populations of cells over only 11-12 days. Furthermore, in our post-MACS samples we are realistically only measuring VSG expression in hundreds to thousands of parasites, many of which were shown by IFA to still be expressing VSG AnTat1.1 (Figure 4.4). Are we generating, and analysing, enough VSG switched cells in order to make conclusions about diversity? We chose to analyse VSG diversity in in vitro populations of cells over approximately 12 days for three primary reasons. Firstly, by repeating the growth phase of the VSG switch assays, our VSGseq results would complement our data from Chapter 3 and provide insights into what the 'VSG

AnTat1.1 negative population' consisted of (as previously discussed). Secondly, it has recently been suggested that VSG hierarchy is decided by the metabolic costs of expressing a given VSG, i.e. parasites will preferentially activate a shorter VSG which is not so energetically unfavourable (Liu et al., 2018). If we were to culture the cells for long periods of time, it could be that cells expressing shorter VSGs outgrow the rest of the population and the variants present earlier in the growth phase are lost from the analysis. Finally, as discussed in Chapter 3, to avoid the loss of rare variants through bottle-necking and the density-dependent differentiation of uninduced pleomorphic populations, the cell culture volume must be 'scaled-up' rather than diluted. To continue in our assays format for longer periods of time would require unrealistically large volumes of HMI-9. Could we then use a chronic mouse model instead? In our experimental set-up, and to answer our questions, no. A chronic infection mouse model would require the trypanosomes to differentiate to stumpy forms, however our inducible monomorphs were observed in Figure 3.5 to replicate to extremely high densities with little to no sign of growth arrest. Therefore, it would be highly unlikely these cells could be grown for long periods of time in any mouse model without risk of significant pathology or host death. Our uninduced differentiation competent reporter cell lines could of course be kept in a chronic mouse model however, the production of stumpy forms would limit the expressed VSG repertoire since they cannot form new variants. This would therefore not allow for a fair comparison between the uninduced and induced conditions. Thus, whilst measuring VSG expression diversity in small populations of cells after 11-12 days does reduce accuracy and not shed light on parasite-host interactions, in the context of our study and asking whether antigen switch rate and slender to stumpy differentiation capacity are mechanistically tethered, we believe our approach is informative.

Only ~5% of the annotated archival VSG genes are functional (Marcello and Barry, 2007). Mosaic VSG gene formation, catalysed by segmental gene conversion, is therefore central to the expansion of the pre-existing VSG repertoire in the later stages of infection (Hall et al., 2013; Marcello and Barry, 2007; Mugnier et al., 2015). Marcello and Barry (2007), who sequenced subcloned VSGs in a manner similar to Hall et al. (2013), identified mosaic VSG cDNA from day 22 of infection

onwards, Hall et al. (2013) from day 16 of infection, and Mugnier et al. (2015) from day 21 of infection. Using the criteria set out by Mugnier et al. (2015) (see Section 4.5) no candidate mosaic VSGs were identified in our datasets. The parasites from which the RNA was isolated had only been cultured in the absence of selection for 12 days and therefore it was highly unlikely that a mosaic VSG would be found, though to the best of our knowledge, no other *in vitro* study has identified the presence of mosaic VSGs. It would be interesting to observe if, following longer periods of unselected growth, mosaic VSGs would occur in an *in vitro* culture or whether, in the absence of immune selection, parasites would preferentially revert back to the expression of previously expressed VSGs once the repertoire of intact VSG genes (estimated to be ~400 for the Lister 427 strain (Cross et al., 2014)) was exhausted.

In summary, VSGseq analysis of pleomorphic and inducibly monomorphic populations of *T. brucei* has revealed vast levels of VSG expression diversity, but the induction of monomorphism was not associated with a reduction in VSG expression diversity compared to the pleomorphic population, or the selection of VSGs distinct from those activated by the pleomorphic population. The conclusions reached from this work corroborate the findings in Chapter 3, where it was first found, through FACS-based VSG switch assays, that VSG switch frequency and stumpy differentiation capacity can be uncoupled.

5 : MEASURING THE VSG SWITCH RATE OF *IN VITRO* GENERATED SELECTED MONOMORPHS

5.1 Introduction

The literature states that monomorphic trypanosomes switch their expressed VSG surface coat at a significantly lower frequency than pleomorphic trypanosomes, suggesting that antigen switch frequency and stumpy differentiation capacity may be linked. Through *in vitro* FACS-based VSG switch assays (Chapter 3) and targeted VSGseq analysis (Chapter 4) however, I have shown that VSG switch frequency and slender to stumpy differentiation capacity can be uncoupled. By adopting an inducible monomorphism model (whereby induction of RNAi in pleomorphic cells against a component of the stumpy formation pathway rendered the population insensitive to SIF), the VSG switch frequency and VSG expression diversity of the same cell lines, in which differentiation capacity was either switched 'on' or 'off', was addressed. VSG switch assays demonstrated that upon the induction of monomorphism in three separate $GFP^{ESpro}AnTat1.1^{ES}$ RNAi lines, each of the cell lines lost their differentiation capacity but continued to switch VSG expression at rates between $6x10^{-4}$ and $2x10^{-3}$ switches/cell/generation. Furthermore, induction of monomorphism similarly did not elicit a change in the proportion of cells switching their expressed VSG by DNA recombination. VSGseq analysis revealed high levels of VSG expression diversity, however the induction of monomorphism was not associated with a reduction in VSG expression diversity, nor the selection of distinct VSG genes.

These results contradicted what has been reported in the literature based on monomorphs generated by serial passage. As such, a second strategy was adopted to verify the conclusions from Chapters 3 and 4. The approach entailed generating populations of selected monomorphs *de novo* through repeated *in vitro* passage of the pleomorphic reporter cell lines used in Chapters 3 and 4 and comparing their VSG switch rates to the respective parental populations using the FACS-based VSG switch assay. Since the same cell line is followed throughout the process of selection, and the switch rates both before and after selection are measured, the results can thus make it very clear whether VSG switch frequency and slender to stumpy differentiation capacity are coupled, or can be selected for independently of one another during long term passage.

5.2 Long term culture of trypanosome populations causes increased resistance to 8-pCPTcAMP *in vitro*

Historically, virulent monomorphic trypanosome lines have been generated by the repeated syringe-passage of parasites through rodents. Turner (1990) estimated that to lose pleomorphism by this method, cells would need to be passaged between mice every 2-3 days between 30 and 50 times. Therefore, to replicate the methods used in these studies for each of our three reporter cell lines, would require between 90 and 150 mice. Virulent cells incapable of differentiation to cell cycle arrested stumpy forms are selected for in the absence of tsetse fly transmission, a condition which is obviously easily attainable in a culture model. Therefore, to reduce the number of rodents used, we chose to generate our selected monomorphs *in vitro*.

To generate monomorphic populations of the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi cell lines used in Chapters 3 and 4, the populations were subjected to repeated *in vitro* passage in culture flasks for a period of 58-73 days (Figure 5.1). To avoid the accumulation of SIF, and therefore cell cycle arrest, pleomorphic bloodstream form (BSF) parasites should be maintained at densities below 10^6 cell/ml (Hirumi and Hirumi, 1989; Vassella et al., 1997a). To select for cells insensitive to a density dependent differentiation signal, the culture densities were gradually increased until the populations could be maintained between densities of 1-3x10^6 parasites/ml. Cells were kept in selective drug medium to prevent loss of expression of both the GFP and RNAi machinery, but RNAi was not induced. To aid in tracing the dynamics of selection, an aliquot of each population was frozen at -80^{o}C on roughly a weekly basis.

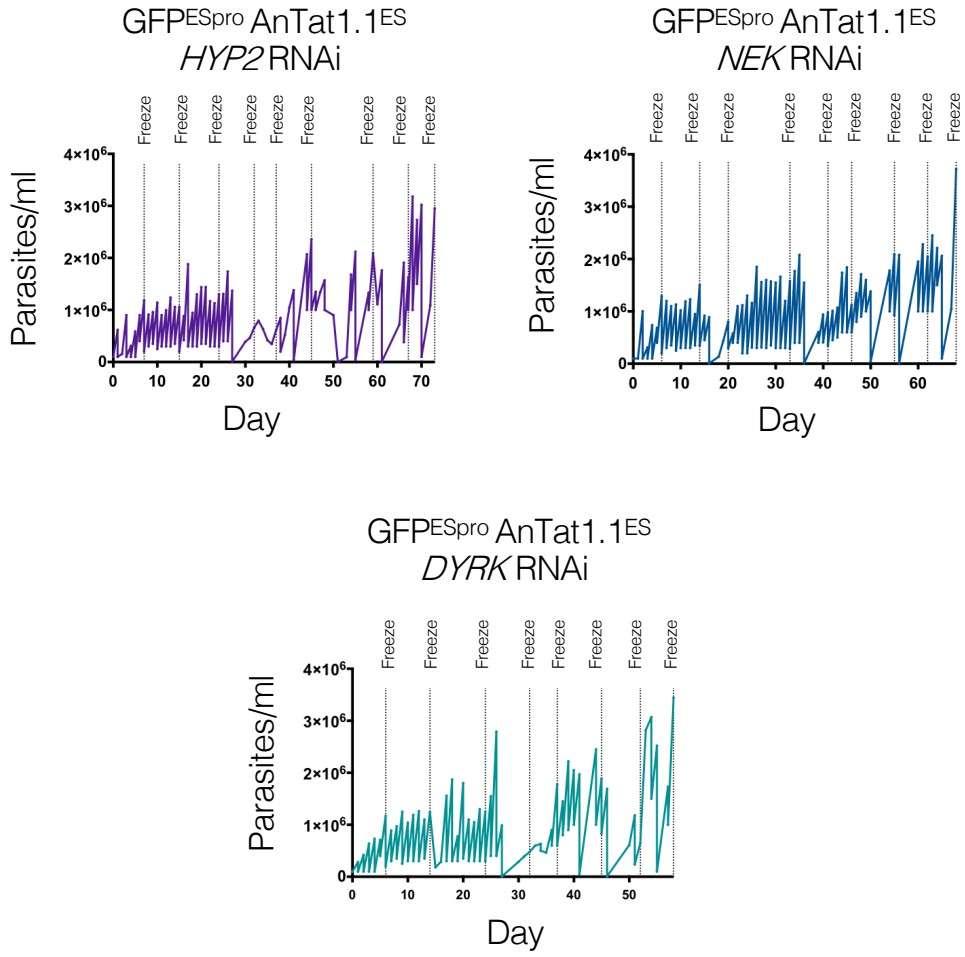


Figure 5.1 *In vitro* selection for monomorphism by repeated passage at high parasite density. Growth profiles for the uninduced GFP^{ESpro}AnTat1.1^{ES} *HYP2* (purple), *NEK* (blue) and *DYRK* (teal) RNAi cell lines. Days when an aliquot of the respective population was frozen are shown by the dotted lines.

After the selection period, all three of the populations displayed noticeably reduced population doubling times and were able to maintain proliferation at markedly higher cell densities compared to the respective 'Start' populations. Rapid proliferation and growth at high densities are typical characteristics of monomorphic cell populations and therefore the final selected populations were tested *in vitro* for their response to incubation with the cell permeable SIF mimic 8-pCPTcAMP (Figure 5.2).

As expected, the uninduced pleomorphic 'Start' populations for all three cell lines rapidly underwent cell cycle arrest upon incubation with 100μM 8-pCPTcAMP. The final selected populations, however, differed in their responses- both to their respective pleomorphic 'Start' populations and to each other. The selected GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi population showed increased resistance to 8-pCPTcAMP compared to the 'Start' population, however its growth had begun to slow by day 4. In contrast, both the selected GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi lines sustained their growth in the presence of 8-pCPTcAMP, growing only marginally slower than selected populations in the absence of 8-pCPTcAMP and equally as well as their respective 'Start' populations in the absence of 8-pCPTcAMP. To confirm that this decrease in sensitivity to 8-pCPTcAMP was not the result of the loss of regulation of the RNAi machinery during the selection process, the selected populations were induced with doxycycline for three days and the respective levels of *HYP2*, *NEK* or *DYRK* transcripts probed by Northern Blot and compared to an uninduced and wild type (WT) control (Figure 5.2 insets). In all three selected populations, RNAi of the quorum-sensing components remained inducible and therefore the increased resistance to 8-pCPTcAMP could not be explained by leaky expression of T7 RNA polymerase but instead was more likely due to a true loss of pleomorphism.

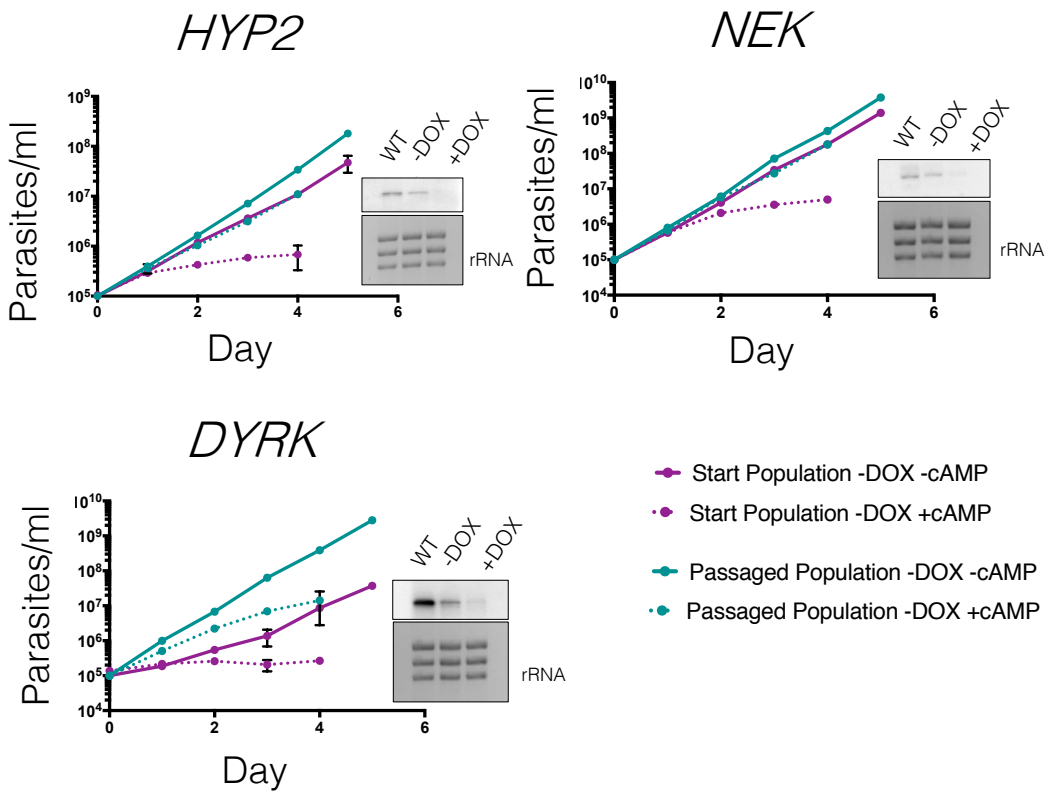


Figure 5.2 8-pCPTcAMP response cumulative growth profiles of the pleomorphic 'Start' and selected GFP^{ESpro}AnTat1.1^{ES} populations. Treatment with 8-pCPTcAMP caused the pleomorphic parental 'Start' populations to arrest. Passaged populations showed varying degrees of resistance to the compound. The 8-pCPTcAMP was added on day 1. The gene targeted by the RNAi machinery is denoted above the respective graph. Data represents the mean±S.D (n=3). **Inset** Northern blot detection of *HYP2*, *NEK* and *DYRK* transcripts in the selected populations. A pleomorphic *T. brucei* AnTat1.1 90:13 line was used as the WT control. The RNAi machinery remained inducible and therefore knock down of a stumpy formation gene could not explain the acquired 8-pCPTcAMP resistance. rRNA was used as a loading control.

To explore if the acquisition of 8-pCPTcAMP resistance proceeded as a gradual process over the period of *in vitro* selection or whether the phenotype occurred suddenly, each of the frozen 'Intermediate' aliquots were thawed and tested for their 8-pCPTcAMP resistance (Figure 5.3). Remarkably, each of the three GFP^{ESpro}AnTat1.1^{ES} RNAi lines had already increased in resistance to 8-pCPTcAMP as early as 7 days into the selection period, when the cultures were still being

maintained at densities below 10^6 cells/ml. This modest rise in resistance gradually increased with continued passage, although an obvious jump in resistance approximately midway through the selection process was observed in each of the lines (Figure 5.3 (a), (b) and (c), top right panels). This did not occur simultaneously between the three populations and nor did it always immediately follow a large population bottleneck event. In the presence of 8-pCPTcAMP, the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi selected populations grew at 10-15% of the rate of the untreated selected population. The GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs, however, exhibited much greater resistance, growing to 60% of that of the untreated cells.

Since a decrease in generation time was also observed in the selected populations, the cumulative growth of the final and each of 'Intermediate' populations in -8-pCPTcAMP conditions was quantified and compared with the respective starting population (Figure 5.3 (a), (b) and (c), bottom right panels). The large shift in 8-pCPTcAMP resistance always preceded the drop in generation time (visualised as an increase in cumulative growth). However, the increases in 8-pCPTcAMP resistance and decreases in generation time did not precisely track each other, and nor was high 8-pCPTcAMP resistance always an indicator of lower generation time (and vice versa). All three cell lines showed a gradual increase in growth rate during the *in vitro* passage selection. Interestingly, although the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi lines showed similar levels of resistance to 8-pCPTcAMP, the kinetics of their changes in relative growth compared to the 'Start' populations were very different. Whilst the GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi population showed a slow and moderate rise to ~2-fold greater cumulative growth, the GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi population instead remained relatively stable until a large shift to ~3-fold greater cumulative growth in the latter stages of selection. If this drop in generation time was coupled to a plateau in 8-pCPTcAMP resistance, this could explain why the population appears to lose 8-pCPTcAMP resistance in the final weeks of selection. The inconsistency between the dynamics of 8-pCPTcAMP and growth rate selection between the three lines indicates that 8-pCPTcAMP resistance and rapid generation time were likely selected for independently of each other.

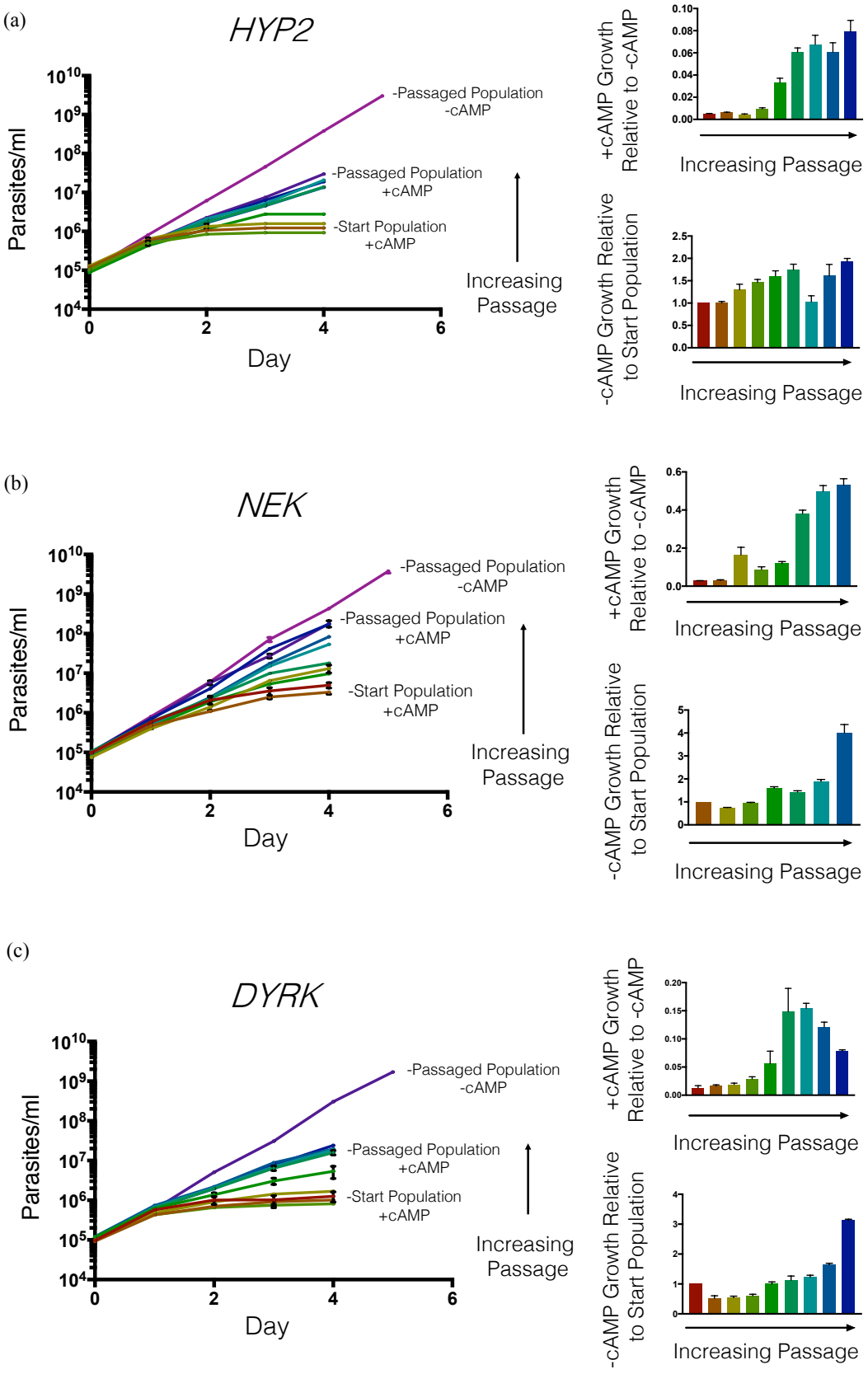

(a)
HYP2
Parasites/ml
Day
-Passaged Population
-cAMP
-Passaged Population
+cAMP
-Start Population
+cAMP
Increasing
Passage
+cAMP Growth
Relative to -cAMP
Increasing Passage
-cAMP Growth Relative
to Start Population
Increasing Passage
(b)
NEK
Parasites/ml
Day
-Passaged Population
-cAMP
-Passaged Population
+cAMP
-Start Population
+cAMP
Increasing
Passage
+cAMP Growth
Relative to -cAMP
Increasing Passage
-cAMP Growth Relative
to Start Population
Increasing Passage
(c)
DYRK
Parasites/ml
Day
-Passaged Population
-cAMP
-Passaged Population
+cAMP
-Start Population
+cAMP
Increasing
Passage
+cAMP Growth
Relative to -cAMP
Increasing Passage
-cAMP Growth Relative
to Start Population
Increasing Passage

Figure 5.3 Changes in 8-pCPTcAMP resistance and growth rates over a 10-week *in vitro* period of passage. Central figures. The three large central graphs plot the *in vitro* cumulative growth profiles of the **(a)** *HYP2* **(b)** *NEK* and **(c)** *DYRK* RNAi GFP^{ESpro}AnTat1.1^{ES} populations in response to 8-pCPTcAMP exposure with increasing passage time. The cumulative growth of the final selected populations in the absence of 8-pCPTcAMP are shown for reference. **Top right figures** Relative cumulative growth of the 'Start', 'End' and 'Intermediate' populations in the presence of 8-pCPTcAMP compared to the cumulative growth of the respective populations in the absence of the compound. With increasing passage, resistance to 8-pCPTcAMP increased. The *HYP2* dataset is missing the pleomorphic 'Start' population and therefore values were normalised to the + 1 week population. The *NEK* dataset is missing the + 1 week population. The data represents the mean ± S.D. (n=3). **Bottom right figures** Relative cumulative growth of the 'Intermediate' and final populations in the absence of 8-pCPTcAMP compared to the parental population. With increasing passage number, the generation time decreased. The data represents the mean ± S.D. (n=3). The *NEK* dataset is missing the 'Start' and + 1 week population data and therefore the values were normalised to the +2 week values.

5.3 A selected monomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population does not respond to SIF *in vivo*

To investigate if the increase in 8-pCPTcAMP resistance corresponded to *in vivo* resistance to SIF, duplicate mouse infections with the 'Start' and selected populations for each of the three GFP^{ESpro}AnTat1.1^{ES} RNAi cell lines were initiated. Parasitaemia, morphology and cell-cycle status were monitored over the course of infection and PAD1 expression was assessed from trypanosomes purified from the blood on the final day of infection. Since monomorphic trypanosomes would be predicted to replicate without restraint in the mouse model, the infections with the selected trypanosome infections were initiated with less cells to allow time for the pleomorphic 'Start' populations to arrest and differentiate in response to the accumulation of SIF. The selected GFP^{ESpro}AnTat1.1^{ES} NEK RNAi population displayed the highest resistance to 8-pCPTcAMP *in vitro*. *In vivo*, the selected GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cells continued to proliferate over 5 days, whilst the parental pleomorphic cells committed to differentiation from day 4 (Figure 5.4 a).

The parental pleomorphic cells in the blood on day 5 of infection had transformed morphologically to stumpy forms with pronounced undulating membranes, however the selected populations, despite being at a far greater parasitaemia, retained the typical slender morphology associated with proliferative bloodstream forms (Figure 5.4 b). Consistent with these observations, the proportion of non-selected pleomorphic cells with a 1K1N configuration increased as the cells differentiated and the parasitaemia plateaued. In contrast, the proportion of selected cells in 1K1N remained constant throughout the infection, indicating an absence of G1/G0 arrest (Figure 5.4 c). Western blot analysis furthermore revealed PAD1 expression to be greatly diminished in the selected populations on day 5 of infection compared to the pleomorphic 'Start' populations (Figure 5.4 d). Taken together, this data suggests that the sustained *in vitro* passage of the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cell line had selected for a population of cells that had lost their differentiation capacity. As such, these cells will now be referred to as 'selected monomorphs'.

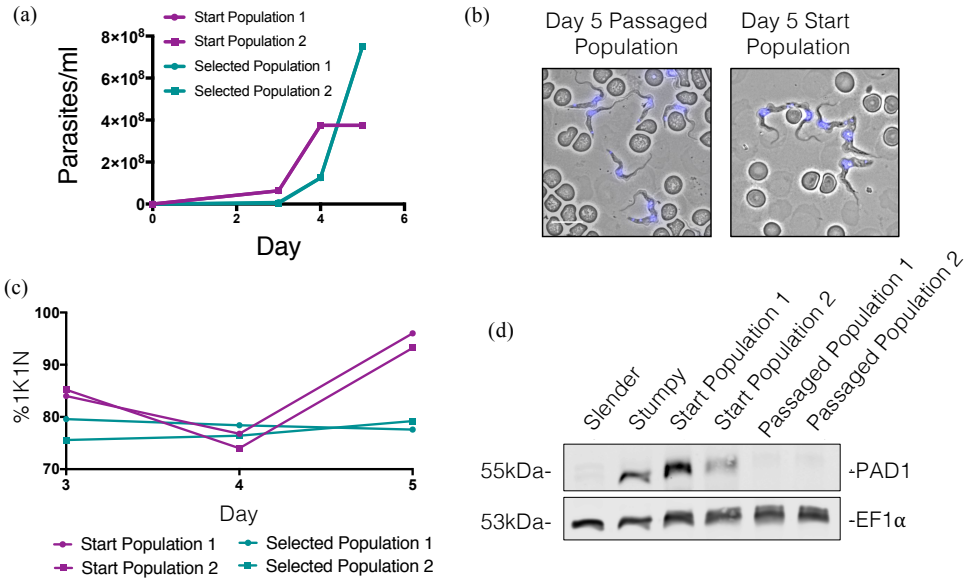


Figure 5.4 GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs do not respond to SIF *in vivo*. (a) *In vivo* growth of *NEK* RNAi GFP^{ESpro}AnTat1.1^{ES} selected monomorphs compared to the pleomorphic parental population. Selected monomorphs remained proliferative, whilst the parental cells arrested in response to SIF. **(b)** Morphological analysis of methanol-fixed cells isolated on the final day of infection. Parental cells (right) displayed

the typical morphological characteristics of stumpy cells, such as a shortened free-flagellum, pronounced undulating membrane and an expanded flagellar pocket, whilst selected monomorphs (left) remained largely long and slender. Nuclear and kinetoplast DNA was visualised by DAPI staining. Bar=10μm. **(c)** Cell-cycle analysis of cells in the blood on days 3, 4 and 5 of infection. Pleomorphic cells (purple) arrested in G1/G0 as they differentiated, seen as an accumulation of 1K1N cells between days 4 and 5. The proportion of cells in 1K1N within the selected monomorph population (teal) remained constant. 250 cells were counted for each individual infection. **(d)** Western blot detection of PAD1. At high parasitaemia, selected monomorphs did not express PAD1 on their surface. Slender and stumpy *T. brucei* AnTat1.1 90:13 WT cells served as controls. Detection of EF1α was performed as a loading control.

The GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi selected population displayed moderate levels of 8-pCPTcAMP resistance *in vitro* (Figure 5.2). In the mouse model, the selected population's parasitaemia gradually ascended without indication of plateau (Figure 5.5 a), consistent with the *in vitro* growth rate observations in Figure 5.3. Selected cells isolated on day 6 were morphologically slender and did not express the stumpy specific protein PAD1 (Figure 5.5 b, d). Fluorescence microscopy showed that, at every timepoint, a smaller proportion of the selected population was in G1/G0 compared with the parental population. However, as the passaged populations' parasitaemias increased, there was nevertheless a gradual accumulation of cells in 1K1N (Figure 5.5 c). Therefore, it appears that the GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi passaged population is not entirely monomorphic and instead has a diminished response to SIF, exemplified by its slower accumulation in G1/G0.

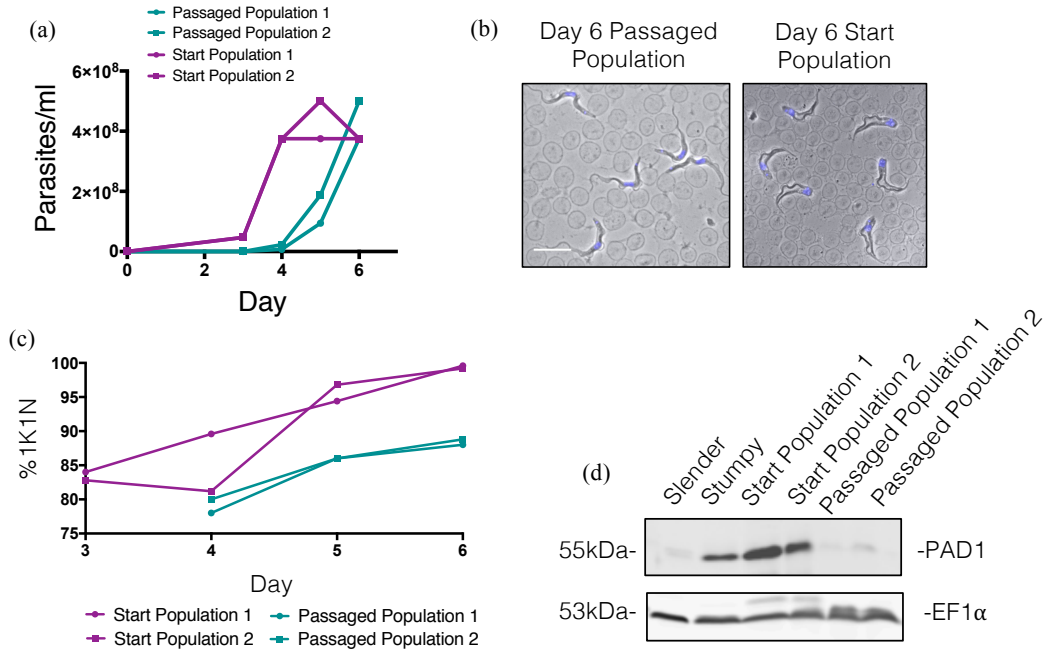


Figure 5.5 The selected GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi population has a diminished response to SIF *in vivo*. (a) *In vivo* growth of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi selected population compared to the pleomorphic parental cells. The selected population's parasitaemia continued to ascend over the infection time course. The pleomorphic parental cells arrested in response to SIF. **(b)** Morphological analysis of cells isolated on the final day of infection. The pleomorphic parental cells became morphologically stumpy (right), whilst the selected cells (left) remained largely long and slender. Nuclear and kinetoplast DNA was visualised by DAPI staining. Bar=10µm. **(c)** Cell-cycle analysis of cells in the blood on days 3-6 (parental population) and days 4-6 (passaged population) of infection. Pleomorphic cells (purple) arrested in G1/G0 as they differentiated, seen as an accumulation of 1K1N cells. The proportion of replicative cells was always greater in the selected population throughout the infection time course, however an accumulation of 1K1N cells as the parasitaemia increased was nevertheless observed. 250 cells were counted for each individual infection. **(d)** Western blot detection of PAD1. The passaged population did not express PAD1 on its surface on the final day of infection. Slender and stumpy *T. brucei* AnTat1.1 90:13 WT cells served as controls. Detection of EF1α was performed as a loading control.

After extensive *in vitro* passage, the GFP^{ESpro}AnTat1.1^{ES} DYRK RNAi selected population displayed levels of 8-pCPTcAMP resistance similar to that of the

GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi selected population. Despite rapid *in vitro* proliferation, the GFP^{ESpro}AnTat1.1^{ES} DYRK RNAi selected cells grew relatively slowly over the first few days of infection, before the parasitaemia finally ascended rapidly from day 5 onwards (Figure 5.6 a). Akin to the response of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi passaged population, the selected GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi cells maintained their slender morphology, even at high parasitaemia (Figure 5.6 b) and a greater proportion of the population remained proliferative (Figure 5.6c). However, again, an accumulation of cells in 1K1N over the infection time course was observed. Since at no point did the parasitaemia of the selected population surpass that of the parental line, it cannot be ruled out that the GFP^{ESpro}AnTat1.1^{ES} DYRK RNAi passaged population may have been observed to morphologically transform and express PAD1 if a later timepoint had been taken.

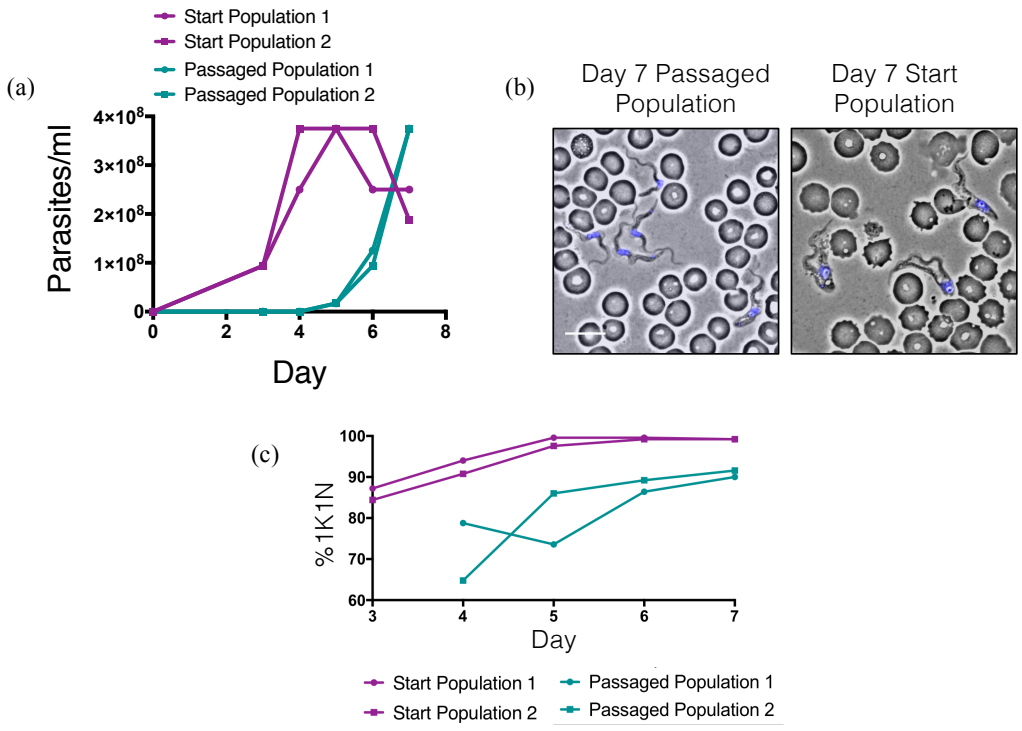


Figure 5.6 The selected GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi population has a low level of resistance to SIF *in vivo*. (a) *In vivo* growth of the GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi selected population compared to the pleomorphic parental cells. The selected population's early growth was delayed, however later continued to ascend as the parental cells

differentiated to stumpy forms. **(b)** Morphological analysis of cells isolated on the final day of infection. Parental cells became morphologically stumpy (right), whilst selected cells (left) remained largely long and slender. Nuclear and kinetoplast DNA was visualised by DAPI staining. Bar=10μm. **(c)** Cell-cycle analysis of cells in the blood on days 3-7 (parental population) and days 4-7 (passaged population) of infection. Pleomorphic cells (purple) arrested in G1/G0 as they differentiated, seen as an accumulation of 1K1N cells (solid line). The proportion of replicative cells was greater in the selected population throughout the infection time course, however a large accumulation of 1K1N cells as the parasitaemia increased was nevertheless observed and by day 7 almost 90% of the selected population had accumulated in 1K1N. 250 cells were counted for each individual infection.

5.4 Selected monomorphs do not exhibit a lower VSG switch rate compared to the parental pleomorphic population

Induction of monomorphism in the GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi lines did not significantly reduce the VSG switch rate, change the mechanism of VSG switching, nor decrease the diversity of expressed VSGs as demonstrated by *in vitro* FACS-based VSG switch assays and VSGseq analysis (Chapters 3 and 4). This therefore provides strong evidence that stumpy differentiation capacity and VSG switch frequency can be uncoupled. Over the 10 week period of *in vitro* passage, varying degrees of monomorphism were selected for in the 3 populations- the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs demonstrated complete resistance to the accumulation of SIF *in vivo* (Figure 5.4) but whilst the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi lines did not transform morphologically to stumpy forms or express PAD1 on their surface, they did accumulate in G1/G0 over an infection time course (Figures 5.5 and 5.6). To investigate if a reduced VSG switch frequency was simultaneously selected alongside the loss of stumpy differentiation capacity, the VSG switch rates of the passaged GFP^{ESpro}AnTat1.1^{ES} *HYP2*, *NEK* and *DYRK* RNAi populations were tested using the *in vitro* FACS-based VSG switch assays (as described in Section 3.6) and compared to the respective pleomorphic parental population (Figure 5.7a). As would be expected for populations

with increased rates of proliferation, the switch assays measuring the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi passaged cells' switch rates lasted significantly shorter than those for the parental pleomorphs (9 days vs 10.3 $\pm$ 0.5 days, p=0.0015 Welch's t-test and 7 days vs 9 $\pm$ 1.1 days, p=0.0066 Welch's t-test, respectively). However, the mean number of generations undergone in the pleomorph and selected monomorph GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *NEK* RNAi switch assays (26.8 $\pm$ 0.6 vs 26.8 $\pm$ 0.7 and 26.4 $\pm$ 0.7 vs 26.9 $\pm$ 0.4, respectively) were not statistically different (p>0.05). In the GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi pleomorph and selected monomorph switch assays, there was no significant difference in either the mean number of days (9 $\pm$ 1.1 vs 9) or the mean number of generations (27 $\pm$ 0.9 vs 27.5 $\pm$ 0.8) that the assays were run for (p>0.05). Statistical analysis found no significant difference in the mean number of VSG switches/cell/generation calculated for the pleomorphic parental GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi populations compared to the values calculated for the uninduced lines in Figure 3.9 of Chapter 3 ($6.67x10^{-4}$ $\pm$ 0.0005 vs $6.97x10^{-3}$ $\pm$ 0.0004 and $1.43x10^{-3}$ $\pm$ 0.0014 vs $2.01x10^{-3}$ $\pm$ 0.002 switches/cell/generation, respectively). In this set of experiments, the VSG switch rate of the GFP^{ESpro}AnTat1.1^{ES} NEK RNAi parental pleomorphs was measured as significantly lower than the VSG switch rate calculated for the uninduced pleomorphs in Chapter 3 ($3.1x10^{-4}$ $\pm$ 0.00009 vs $6.31x10^{-4}$ $\pm$ 0.0002 switches/cell/generation.

The results of the selected monomorph VSG switch assays clearly show that in all three lines the VSG switch rate had not dropped to the published estimates for monomorphic cells lines, despite the loss of pleomorphism (Figure 5.7b). The GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi selected monomorphs and GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi passaged population each marginally increased their mean switch rates to $1.39x10^{-3}$ $\pm$ 0.0006 and $2x10^{-3}$ $\pm$ 0.0009 switches/cell/generation, respectively, whilst the switch rate of the *NEK* RNAi passaged population decreased to $2.88x10^{-4}$ $\pm$ 0.0001 switches/cell/generation. None of these changes, however, were significant (p>0.05).

Monomorphic cells are reported to preferentially switch their expressed VSG by *in situ* switches of transcription (Liu et al., 1985; McCulloch et al., 1997;

Robinson et al., 1999). To investigate if selection for monomorphism and *in situ* switching occurred concurrently, the VSG AnTat1.1 negative cells were observed for their GFP positivity (Figure 5.7c). In these assays, the pleomorphic GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi lines switched VSG expression by DNA recombination at a similar extent to what was seen previously in the uninduced assays performed in Figure 3.9 (49.93 ± 25.01% vs 44.67 ± 31.43% and 75.62 ± 20.3% vs 84.63 ± 11.16%, respectively). The pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi line under these experimental conditions however, switched significantly less by DNA recombination (25.02 ± 16.46% vs 52.7 ± 24.53%, p=0.045 t-test). These results were all underpinned by high variability, however. Figure 5.7c shows that, during the 10 week period of passage, the selection of monomorphism was not directly accompanied by a selection for increased *in situ* VSG switches. Conversely, switching by DNA recombination appeared to increase in the monomorphic GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi lines compared to the pleomorphic 'Start' populations (54.63 ± 37.9% vs 49.93 ± 31.4% and 91.78 ± 1.8% vs 75.62 ± 20.3%, respectively). None of the observed changes were statistically significant, however it must be noted that the results for the GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi selected monomorphs, in particular, were characterised by high rates of variability.

The literature states that pleomorphic cells switch their expressed surface VSG at a rate of ~10^{-4} to 10^{-2} switches/cell/generation and monomorphic cells at ~10^{-7} to 10^{-5} switches/cell/generation (Turner, 1997). In this set of experiments, the VSG switch rate of three cell lines before and after a ten-week period of *in vitro* selection were measured. Whilst the ability of the cells to effectively differentiate to transmission-competent stumpy forms was lost during the period of selection, the VSG switch rate remained consistently high, thus demonstrating that VSG switch frequency and stumpy differentiation capacity can be uncoupled.

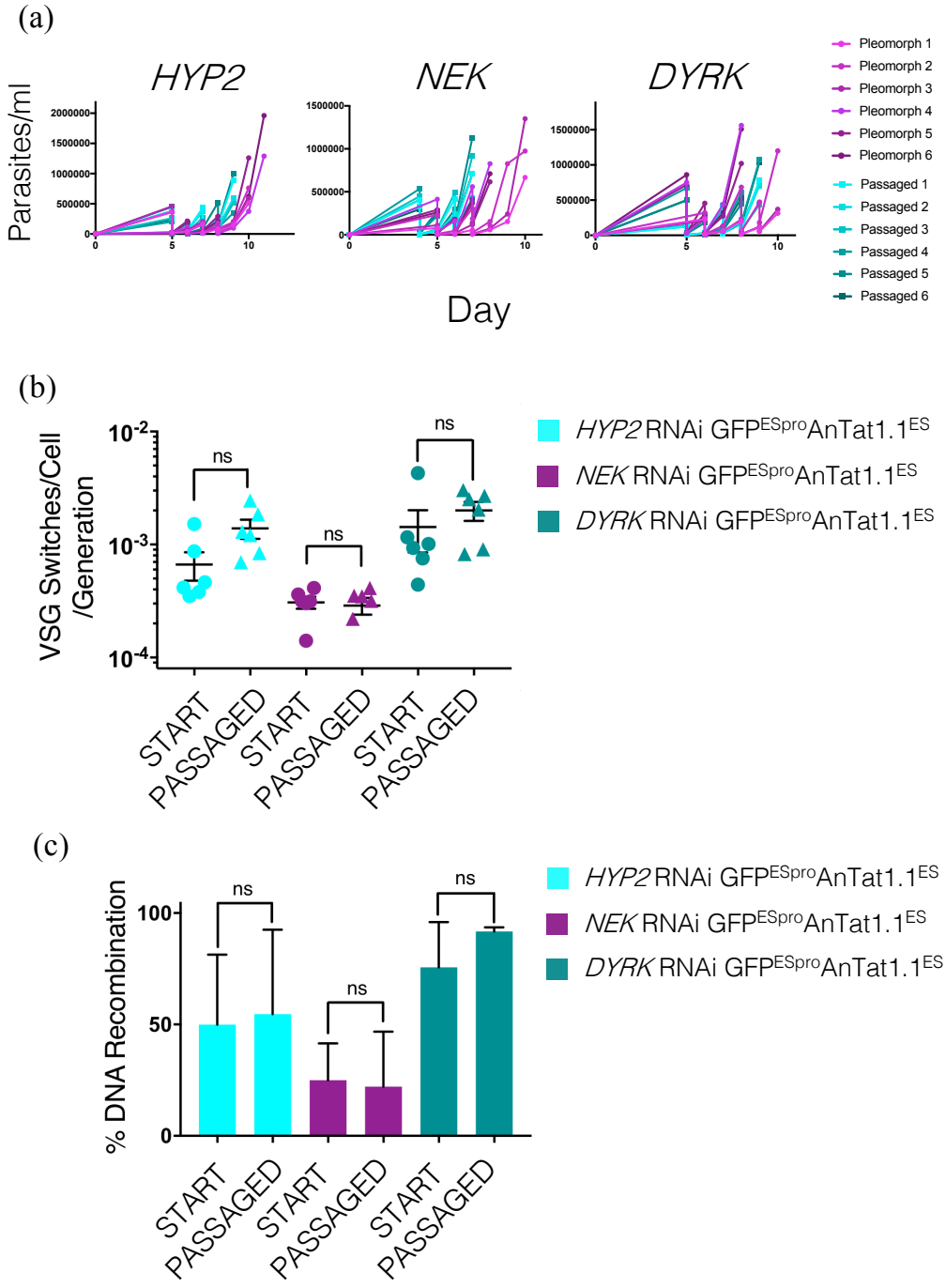


Figure 5.7 Monomorphism can be selected for independently of VSG switch rate (a) VSG switch assay growth of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* (left), *NEK* (middle) and *DYRK* (right) RNAi pleomorphs and selected monomorphs. The parental pleomorphic populations are shaded in plum and the passaged populations in teal. The assays were performed twice in triplicate (n=6), except for the pleomorphic *DYRK* RNAi experiment where one assay was performed in duplicate (n=5). **(b)** The calculated VSG switches/cell/generation for the

parental and selected GFP^{ESpro}AnTat1.1^{ES} *HYP2* (blue), *NEK* (plum) and *DYRK* (teal) RNAi lines. There was no significant difference in VSG switches/cell/generation after selection for monomorphism, p>0.05, t test. **(c)** The percentage of switched GFP^{ESpro}AnTat1.1^{ES} *HYP2* (blue), *NEK* (plum) and *DYRK* (teal) RNAi cells that had switched their expressed VSG by DNA recombination (i.e. VSG AnTat1.1 negative/ GFP positive). Selection for monomorphism was not associated with a significant reduction in DNA recombination based switching (p>0.05, t-test). n=6 except for: passaged GFP^{ESpro}AnTat1.1^{ES} *HYP2* RNAi population (n=5); pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* population (n=5); passaged GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population (n=4); and pleomorphic GFP^{ESpro}AnTat1.1^{ES} *DYRK* RNAi population (n=5).

5.5 gDNA and RNA sequencing of pleomorphic and monomorphic *NEK* RNAi GFP^{ESpro}AnTat1.1^{ES} populations

Evolve and resequence (E & R) experiments combine experimental evolution and next generation sequencing to follow molecular evolution at a genome-wide scale in real time. By sequencing populations exposed to controlled selective conditions, researchers can explore the molecular components of adaptation through the identification of single nucleotide polymorphisms (SNPs) and causative genes (Long et al., 2015). Sustained *in vitro* passage at high parasite density caused varying degrees of monomorphism in the three selected populations. This was most apparent in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi line, where a population of selected monomorphs that remained highly proliferative, morphologically slender and did not express the stumpy marker PAD1on their surface in the presence of SIF were selected (Figure 5.4). Application of the E & R approach to understand the loss of pleomorphism upon laboratory adaptation of trypanosome populations could provide us with important insights into the process and dynamics of laboratory adaptation and moreover, identify novel molecular components regulating the hierarchical quorum-sensing differentiation pathway. Therefore, the pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi parental population, monomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected population, and two ‘Intermediate’ samples

were chosen for genomic DNA and total RNA deep sequencing with the aim of identifying novel regulators of slender to stumpy differentiation (Figure 5.8a).

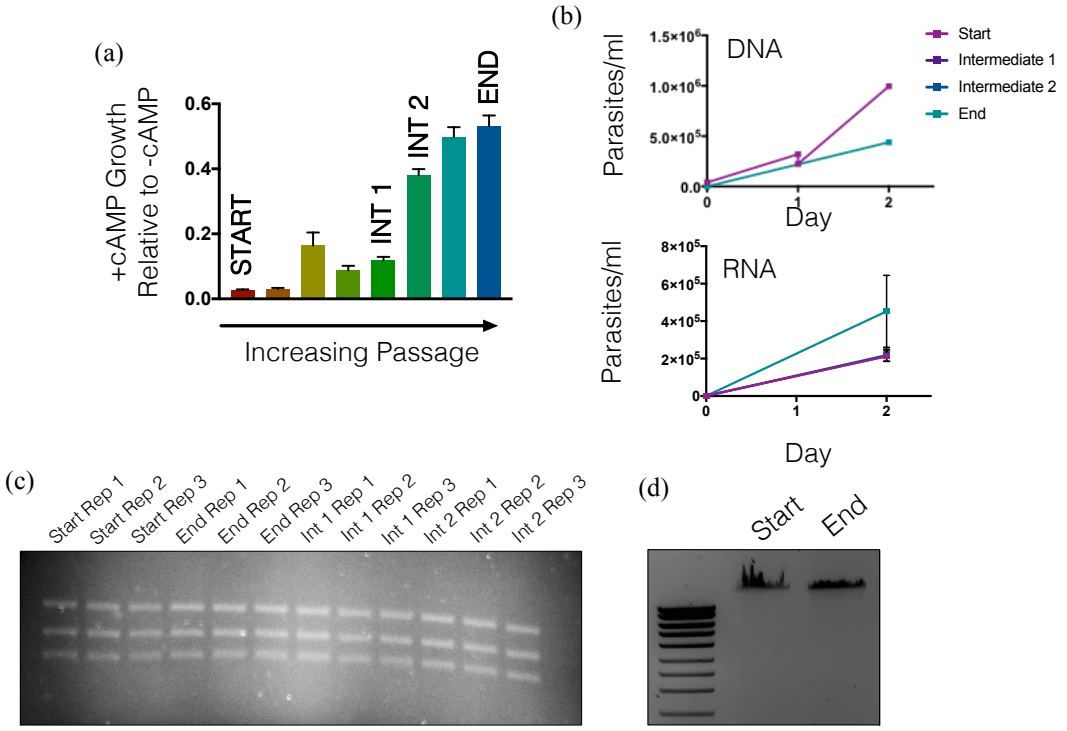


Figure 5.8 Preparation of material for whole genome and transcriptome sequencing of the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs. (a) Total population DNA and RNA from the highlighted time points were sent for sequencing. For genome sequencing, only the 'Start' (pleomorphic) and 'End' (selected monomorph) samples were processed. All four samples had their transcriptomes sequenced. 'Intermediate' sample 1 and 'Intermediate' sample 2 were selected for sequencing since they flanked the marked increase in 8-pCPTcAMP resistance and it was hoped that the changes directly influencing resistance might be obvious in their analysis. **(b)** For each timepoint, triplicate RNA preparations were sequenced and a single gDNA sample. For the triplicate RNA growth curves, the data represents the mean ± S.D. (n=3). Parasites were carefully maintained *in vitro* in logarithmic phase and at densities <10^6 cells/ml before DNA and RNA was isolated. **(c)** The RNA samples were visualised as three tight bands on an ethidium bromide stained formaldehyde gel. 200ng of RNA were loaded per sample. **(d)** The extracted DNA had a high molecular weight and was not degraded or contaminated with impurities. 500ng of DNA were loaded per sample.

Genetic material was prepared from populations of cells that had been expanded *in vitro*. Great care was taken to maintain the densities of the populations from which RNA was extracted under 10^6cells/ml so that any observed changes were not the result of a density dependent response of the pleomorphic population (Figure 5.8b). RNA was extracted from triplicate populations for all four timepoints, whilst single gDNA samples were prepared from the starting pleomorphic and final selected populations. To verify the quality of the prepared material, 200ng of RNA (Figure 5.8c) or 500ng gDNA (Figure 5.8d) were loaded on an agarose gel and visualised with ethidium bromide staining. The extracted RNA was visualised as three tight, and therefore nondegraded, bands and a single, tight, high molecular weight band furthermore signified that the prepared gDNA was also of high quality. Final quality control and the DNA/RNA sequencing services were provided by BGI Tech Solutions, Hong Kong.

5.5.1 Transcriptome analysis of selected monomorphs

Quality control of the raw transcriptome data, read alignments to the *T. b. brucei* TREU927 reference genome and groupwise comparisons of 'reads per kilobase of transcript per million mapped reads' (RPKM)-normalised datasets were kindly performed by Dr Alasdair Ivens (Edinburgh, UK) (see Section 2.18). The initial transcriptome analysis focused on the comparison of the selected monomorph ('End') population relative to the pleomorphic parental ('Start') population. When an adjusted p-value cut off of $p \leq 0.05$ was applied to the dataset, 950 transcripts were found to be significantly less abundant in the selected monomorphs relative to the parental pleomorphs and 578 transcripts significantly more abundant (Figure 5.9, red dots). The vast majority of these changes in abundance, however, were subtle. A Log2FC value of +1 shows that a transcript has increased 2-fold in abundance relative to the comparator, whilst a Log2FC value of -1 indicates that the abundance of the transcript has halved. The highest Log2FC value generated in the 'End' dataset relative to the 'Start' dataset (once pseudogene, putative VSG and 'unlikely hypothetical' transcripts had been filtered from the analysis) was +0.73 (equivalent to less than 2-fold change), whilst the lowest was -1.81 (equivalent to roughly only a

third less transcript). To remove the transcripts showing least changes in abundance, Log2FC boundaries were applied to the upregulated and downregulated gene lists, respectively, such that transcripts showing less up- or down-regulation than these values would not be considered for further analysis. The upregulated transcripts showed the least changes in abundance, and furthermore encoded no realistic regulators of slender to stumpy differentiation. Thus, a Log2FC cut off of +0.5 was applied to the dataset (Figure 5.9, top left). A Log2FC limit of -0.5 would have produced an unmanageably large downregulated dataset (Figure 5.9, bottom right), and therefore a cut off of -0.6 was applied to the downregulated dataset. Imposing this filtration criteria generated shortlists of 11 genes that were significantly upregulated in the selected monomorph population relative to the pleomorphic parental population, and 68 genes that were significantly downregulated. The genes in these groups alone will be studied and referred to for the remainder of this chapter.

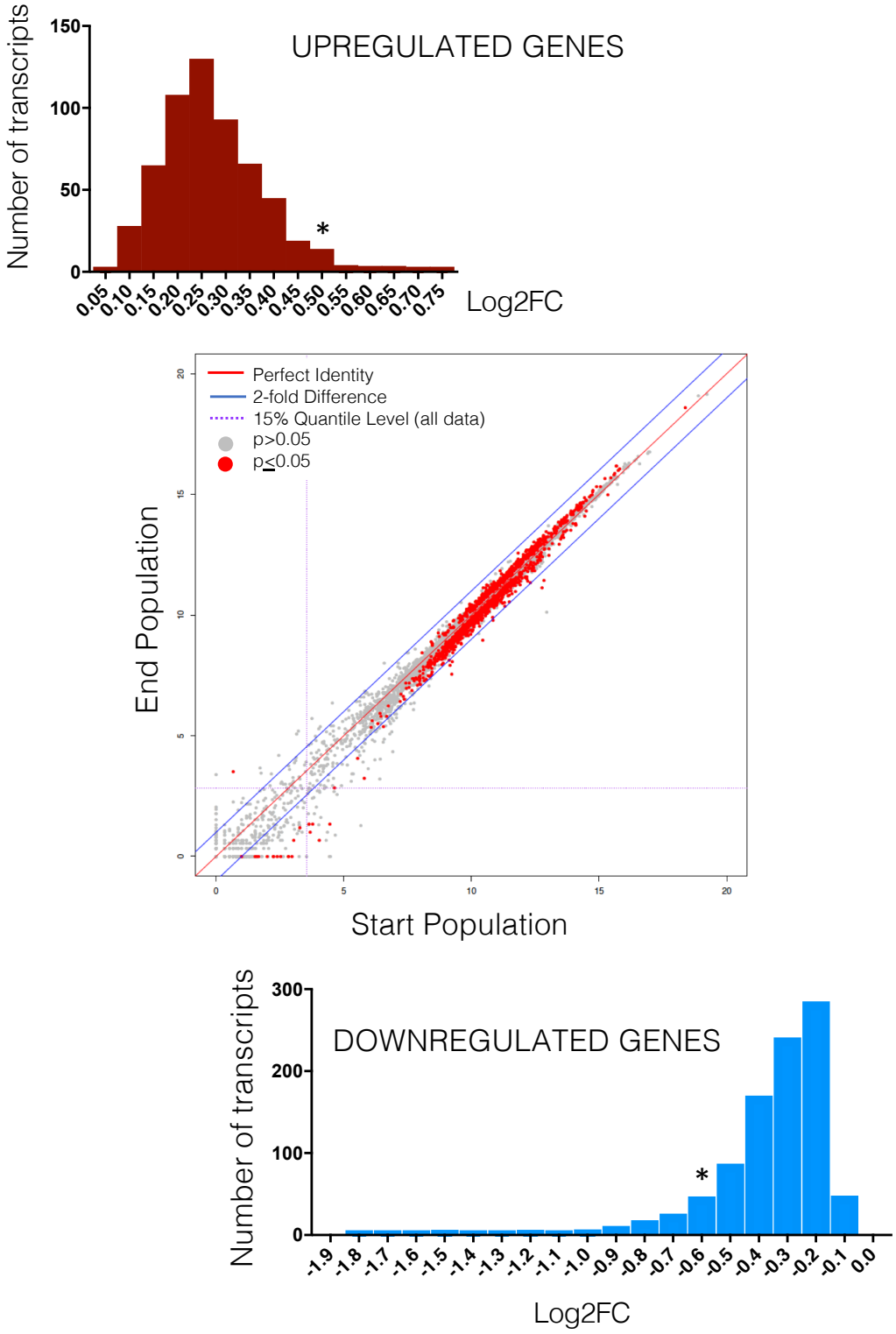


Figure 5.9 GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorph gene expression relative to the pleomorphic parental population. Main panel: Each transcript is denoted by a dot, grey when the p-value was >0.05 and red when the p-value was ≤0.05. The red

horizontal line shows where gene expression was identical between the two populations and the two blue lines represent two-fold increases or decreases in expression. Before pseudogenes and VSG genes were removed from the analysis, 950 genes were recognised as being downregulated (16 over 2-fold) and 578 upregulated (1 over 2-fold). **Top Left:** Transcripts upregulated in the selected monomorphs compared to the parental pleomorphs (pseudogenes and VSG genes were removed from the list). The asterisk denotes the cut off for the upregulated genes shortlist **Bottom Right:** Transcripts downregulated in the selected monomorphs compared to the parental pleomorphs (pseudogenes and VSG gene were removed from the list). The asterisk denotes the cut-off for the downregulated genes shortlist.

5.5.1.1 Selected monomorphs downregulate genes associated with the stumpy and insect forms of the parasite

68 transcripts that were significantly less abundant in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorph population compared to the parental pleomorph population were chosen for analysis. The Log2FC values for the top 9 most decreased transcripts ranged between -1.81 and -1, whilst the remainder extended from -0.97 to -0.61 (Appendix C). The majority of the genes identified (58%) encoded hypothetical proteins, whilst cell surface transporters and receptors were the next most downregulated group (Figure 5.10). Interestingly, 5.5% of the 68 shortlisted transcripts encoded zinc finger proteins, over half of which belong to the novel RNA-binding CCCH (ZC3H) class. Such a finding may be of great significance since the ZC3H class of zinc finger proteins are implicated in the post transcriptional control of gene expression through the binding to, and stabilisation or destabilisation of, mRNA transcripts (Ling et al., 2011). One other downregulated RNA-binding protein, RBP32, was identified in the analysis. A number of protein kinases have been implicated in the regulation of slender to stumpy differentiation, for e.g. DYRK (Mony et al., 2014), MAPK5 (Domenicali Pfister et al., 2006), TOR4 (Barquilla et al., 2012) and NEK17 (Jones et al., 2014). Consistent with this observation, a number of protein kinases and phosphatases were also associated with a significant decrease in transcript abundance after the loss of pleomorphism,

although none was previously detected in earlier screens for differentiation regulators.

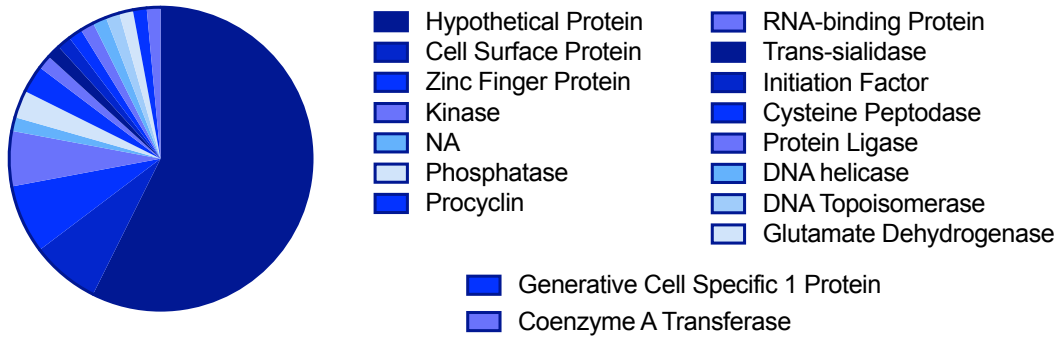


Figure 5.10 Downregulated genes in the selected monomorph population relative to the parental pleomorph population grouped by protein description. Total=68.

The GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs did not respond to SIF *in vivo*, and instead remained as rapidly proliferating slender forms. Consistent with this loss of life cycle competency, many of the 29 annotated downregulated genes had previously been demonstrated as being enriched in stumpy or insect form stages. (Table 5.1).

Table 5.1 Annotated downregulated genes that are associated with the stumpy or insect trypanosome forms.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	GENE FUNCTION
Tb927.6.510	GPEET procyclin	-1.63	0.0003	Major GPI-anchored surface coat protein in procyclic forms (Bütikofer et al., 1997).
Tb927.9.7470	TbNT10	-1.06	0.0003	Purine nucleoside transporter (Spoerri et al., 2007).
Tb927.7.5940	Protein Associated with Differentiation 2 (PAD2)	-0.97	0.0003	Recieves CCA procyclic differentiation signal (Dean et al., 2009).
Tb927.10.10260	EP1 procyclin	-0.96	0.0056	Major GPI-anchored surface coat protein in procyclic forms (Bütikofer et al., 1997).
Tb927.9.11050	4E-interacting protein (4EIP)	-0.78	0.0017	Contributes to tranlsation repression during differentiation to stumpy form (Terrao et al., 2018).
Tb927.7.5930	Protein Associated with Differentiation 1 (PAD1)	-0.67	0.0014	Expressed on surface of stumpy forms (Dean et al., 2009).
Tb927.5.320	Receptor-type adenylate cyclase (GRESAG 4)	-0.64	0.0024	Midgut specific adenyl cyclase. Marker for late procyclic stage (Imhof et al., 2014).
Tb927.11.2690	Succinyl-CoA:3-ketoacid coenzyme A transferase	-0.64	0.0229	Mitochondrial enzyme (Vertommen et al., 2008).
Tb927.7.6830	Trans-sialidase (putative)	-0.64	0.0006	Membrane-bound and expressed on surface. Enable procylic cells to utilise host sialic acid (Engstler et al., 2003).
Tb927.9.5900	Glutamate dehydrogenase	-0.63	0.0029	Mitochondrial enzyme (Mantilla et al., 2017).

The surface proteins PAD1 (Dean et al., 2009) and *Tb*NT10 (Spoerri et al., 2007) are both stumpy markers, although PAD1 is the only protein regularly used in the laboratory as a diagnostic measure for stumpy formation. *Tb*NT10 is the main

P1-type purine transporter of adenosine, guanosine and inosine in procyclic trypanosomes (Al-Salabi et al., 2007). The expression of the gene is developmentally regulated, with the highest expression being found in the stumpy form, followed by procyclic cells and finally, at barely detectable levels in the proliferative slender form (Sanchez et al., 2004). Expression levels seem to be strain specific, however. For instance, in the *T. brucei* AnTat1.1 line, *Tb*NT10 was reported to be expressed at the highest level in the procyclic form of the parasite, although expression was still greatly upregulated in the stumpy form compared to the slender (Spoerri et al., 2007). Being a surface transporter upregulated in the form of the parasite pre-adapted for transmission, *Tb*NT10 has been tested for a role in SIF uptake. A *Tb*NT10 null mutant in the pleomorphic *T. brucei* AnTat1.1 line was viable, and furthermore able to complete a full life cycle *in vivo* however, thus demonstrating that *Tb*NT10 was neither essential nor the receptor of SIF (Spoerri et al., 2007). PAD2 is fundamental to the process of stumpy to procyclic form differentiation within the tsetse midgut. PAD2, a surface-carboxylate transporter, is upregulated and localised to the surface upon cold-shock, sensitising the cell for the perception of the CAA procyclic differentiation signal in the tsetse midgut (Czichos et al., 1986; Dean et al., 2009; Engstler and Boshart, 2004).

4E-interacting protein (4EIP) interacts with the *T. brucei* cap-binding protein EIF4E1 (Terrao et al., 2018). Both proteins have been demonstrated as strong repressors of gene expression when tethered to a reporter mRNA, with 4EIP specifically inhibiting the initiation of translation and promoting mRNA decay (Terrao et al., 2018). BSF and PCF cells deficient in 4EIP display little to no growth phenotype, however upon differentiation to stumpy forms, pleomorphic Δ*4EIP* BSF trypanosomes consistently display elevated protein synthesis compared to arrested WT stumpy cells, as determined by [^{35}S] methionine incorporation into proteins. Furthermore, when these Δ*4EIP* stumpy cells were stimulated to differentiate to procyclic forms by treatment with CCA and incubation in procyclic medium at 27^{o}C, the cells did not accumulate EP procyclin on their surface nor generate a stable proliferative culture (Terrao et al., 2018). Therefore, 4EIP appears to be required for normal BSF to PCF differentiation, specifically in the repression of translation

initiation during stumpy form differentiation. 4EIP-bound mRNAs have comparatively short median half-lives, and are therefore unstable. The bound mRNAs, however, are not enriched for PCF specific transcripts, and nor is any protein motif or family found to be significantly enriched (Terrao et al., 2018).

The remaining transcripts in Table 5.1 are enriched in the procyclic form. GPEET and EP procyclin are the major surface glycoproteins present on the surface of procyclic form cells, whilst succinyl-CoA:3-ketoacid coenzyme A transferase and glutamate dehydrogenase are enzymes involved in mitochondrial metabolism. The receptor-type adenylate cyclase *Tb*927.5.320 was shown by SILAC to be upregulated ~6.25-fold in late compared to early procyclic forms (Imhof et al., 2014). Most receptor-type adenylate cyclases (ACs) belong to the large polymorphic Genes Related to ESAG4 (GRESAG4) family. GRESAGs are distinct from the BSF specific, ES-resident ESAG4 in that the ~65 copies are spread throughout the genome and (most) are expressed throughout the parasite's lifecycle (Alexandre et al., 1996; Berriman et al., 2005; Salmon et al., 2012). The universal signalling molecule cAMP is produced from ATP by ACs. In the BSF parasite, ACs inhibit the host's early innate immune response by producing ~250-fold greater levels of cAMP upon phagocytosis into an acidic environment or TNF-mediated lysis (Salmon et al., 2012). The trypanosome-produced cAMP 'hijacks' the host's cAMP signalling pathways and blocks the synthesis of TNFα by inflammatory cells or macrophage through activation of host protein kinase A (Salmon et al., 2012). In the insect form of the parasite, procyclic-specific AC dimers localise to defined subdomains on the flagellum (Saada et al., 2014). Although the function of many of these is not known, it has been shown that ACP1, ACP4 and ACP4 specifically localise to the tip of the flagellum and that the activity of ACP6 is responsible for the regulation of social motility (Lopez et al., 2015; Saada et al., 2014).

RBP10 is a cytosolic, developmentally regulated RNA-binding protein with one RNA recognition motif (RRM) near its N-terminus (Wurst et al., 2012). Made solely in the slender bloodstream form, the protein promotes the retention of bloodstream form characteristics. Ectopic overexpression of RBP10 in PCF induces the expression of BSF transcripts and requires that cells are grown in BSF medium

for survival (De Pablos et al., 2017; Mugo and Clayton, 2017). Conversely, depletion of RBP10 by RNAi in BSF causes a rapid inhibition of growth that is reversed once cells are placed in PCF growth conditions (De Pablos et al., 2017; Mugo and Clayton, 2017). RBP10 binds to PCF-specific mRNAs with a $UA(U)_6$ motif in their 3' UTR, preventing their translation and furthermore, targeting them for destruction (Mugo and Clayton, 2017). Tandem affinity purification of RBP10 bound mRNAs, and subsequent RNAseq analysis of the bound and unbound fractions, identified 260 mRNAs that were at least 3X more abundant in the bound fraction relative to the unbound. Over 40% of the proteins encoded by these mRNAs had already been quantitated during differentiation, with exactly 50% being more enriched in PCF than BSF. Proteins that showed an enrichment 2-12hrs post stumpy differentiation initiation were clearly enriched in the bound fraction (Mugo and Clayton, 2017). Table 5.1 presents evidence that the selected monomorph population was characterised by a global decrease in stumpy and insect specific transcripts. This evidence is further corroborated when considering the proportion of the significantly downregulated transcripts that were bound by the slender retainer, RBP10. Of the 68 downregulated transcripts in this screen, 31 (45.6%) were identified in the bound fraction as part of the Mugo & Clayton study (2017) (Table 5.2). RBP10 binds PCF transcripts and targets them for destruction. Accordingly, many of the transcripts in Table 5.2, such as GPEET procyclin, ZC3H20 and glutamate dehydrogenase, are already known to be expressed highly in the stumpy and/or procyclic forms. The selected monomorph population has therefore significantly downregulated a large proportion of transcripts required for progression to, and maintenance as, insect form procyclic cells.

Table 5.2 Downregulated transcripts that are bound by RBP10. RBP10 targets procyclic specific transcripts for destruction in the bloodstream form of the parasite (Mugo and Clayton, 2017).

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE
Tb927.6.510	GPEET procyclin	-1.63	0.0003
Tb927.7.2660	ZC3H20 (zinc finger protein)	-1.41	0.0004
Tb927.11.15010	NEK21 (serine/threonine-protein kinase)	-1.00	0.0029
Tb927.5.2260	Hypothetical protein, conserved	-0.93	0.0015
Tb927.9.4560	RBP32	-0.92	0.0004
Tb927.7.6330	Hypothetical protein	-0.90	0.0020
Tb927.11.6890	DNA repair and recombination helicase protein, TbPIF1	-0.90	0.0136
Tb927.6.860	Hypothetical protein	-0.81	0.0033
Tb927.3.1650	Hypothetical protein	-0.81	0.0422
Tb927.11.2410	Hypothetical protein, conserved	-0.80	0.0045
Tb927.4.1000	Hypothetical protein, conserved	-0.79	0.0017
Tb927.10.3360	Hypothetical protein, conserved	-0.77	0.0011
Tb927.10.8050	TFIIF-stimulated CTD (C terminal domain) phosphatase, putative	-0.75	0.0064
Tb927.10.10770	Generative cell specific 1 protein, putative	-0.75	0.0009
Tb927.7.6300	Hypothetical protein, conserved	-0.72	0.0126
Tb927.9.2320	Hypothetical protein, conserved	-0.71	0.0016
Tb927.8.4300	Hypothetical protein	-0.70	0.0306
Tb927.10.12060	Hypothetical protein, conserved	-0.70	0.0069
Tb927.8.5350	Hypothetical protein, conserved	-0.69	0.0069
Tb927.3.2920	Hypothetical protein, conserved	-0.66	0.0126
Tb927.7.520	Hypothetical protein, conserved	-0.65	0.0075
Tb927.11.14070	Protein kinase, putative	-0.65	0.0023
Tb927.6.3880	Hypothetical protein, conserved	-0.64	0.0274
Tb927.4.2410	Hypothetical protein, conserved	-0.64	0.0017
Tb927.5.320	Adenylyl cyclase, putative	-0.64	0.0024
Tb927.9.5900	Glutamate dehydrogenase	-0.63	0.0029
Tb927.11.2450	Hypothetical protein	-0.62	0.0092
Tb927.9.7540	Cysteine peptidase, Clan CA, family C2	-0.62	0.0024
Tb927.10.15610	Zinc finger protein, putative	-0.62	0.0299
Tb927.11.1830	Hypothetical protein, conserved	-0.61	0.0004

The observation that the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorph population downregulated a number of stumpy and insect stage enriched transcripts compared to the parental pleomorphic population was an encouraging validation of the RNAseq data. Where this data's real value lies however, is in the identification the factors that regulate the changes in gene expression, and therefore loss of life cycle competency, in response to long term passage.

The trypanosome genome is organised into multiple long polycistronic transcription units (Berriman et al., 2005). Genes are co-transcribed and coupled 5'trans-splicing and 3'poly-adenylation produces the individual mRNAs (Matthews et al., 1994; Ullu et al., 1993). In most eukaryotes, regulation at the level of transcription governs individual mRNA levels. The trypanosome transcriptional machinery however, lacks the capability to regulate the individual control of transcription initiation and as such the parasites rely unusually heavily on post-transcriptional means of gene expression control, such as mRNA processing through stabilisation or degradation and translational efficiency. RNA-binding proteins typically control differential expression of genes by binding to regulatory AU-rich elements (AREs) in the 3'UTRs of their targets (Mayho et al., 2006). The *T. brucei* genome encodes 49 potential CCCH zinc finger proteins (ZC3Hs) that may be able to bind RNA (Benz et al., 2011; Ouna et al., 2012). ZC3Hs are a novel class of zinc finger protein that bind and either stabilise or destabilise RNA (most other zinc finger family members bind DNA or are involved in protein-protein interactions) (Singh et al., 2014). Three ZC3Hs have already been implicated in the regulation of trypanosome differentiation. *Tb*ZFP1 is procyclic specific and has is involved in kinetoplast repositioning (Hendriks and Matthews, 2005), whilst both *Tb*ZFP2 and *Tb*ZFP3 regulate surface procyclin acquisition and cellular remodelling during BSF to PCF transformation (Hendriks et al., 2001; Paterou et al., 2006). *Tb*ZFP3 may also play a wider role in the developmental events associated with BSF to PCF transformation, demonstrated by a significant enrichment for stumpy specific markers in lists of ZFP3 regulated transcripts (Walrad et al., 2012).

Three ZC3Hs and two putative zinc finger proteins were identified as being significantly downregulated in the selected monomorph population relative to the

pleomorphic parental population (Table 5.3). Indeed, one of the ZC3Hs, ZC3H20, was the fourth most downregulated transcript behind GPEET procyclin and two hypothetical proteins. ZC3H20 is stage-specifically enriched in PCF trypanosomes relative to BSF, and is required for their normal growth (Ling et al., 2011). ZC3H20 was the first ZC3H to be described as stabilising, rather than destabilising, its target mRNAs and was furthermore shown to bind two developmentally-regulated PCF-enriched transcripts, mitochondrial carrier protein 12 (MCP12) and *trans*-sialidase (TS-like E), via regulatory elements in their 3' UTRs (Ling et al., 2011). In addition to MCP12 and TS-like E, another 10 transcripts were found to decrease >2-fold in abundance upon inducible knock down of ZC3H20. Three of these (Table 5.4) are also downregulated in the selected monomorph population. Perhaps the most intriguing of these is *Tb*927.5.4020, which encodes a hypothetical protein and is the third most downregulated transcript. *Tb*927.5.4020 is enriched in the PCF form of the parasite, relative to the BSF, with expression peaking at around 12hrs after the initiation of stumpy to PCF differentiation *in vitro* (Jensen et al., 2009; Queiroz et al., 2009). The protein is included in the *T. brucei* cell surface phylome as a member of Fam50 Clade III (Jackson et al., 2013). Fam50 includes major surface glycoproteins related to the invertebrate to vertebrate transition such as BARP and GARP, two GPI-anchored epimastigote specific surface coat proteins (Jackson et al., 2013). Clade III however, contains surface proteins whose expression is specifically increased in PCF. Despite inclusion in this group, no GPI-anchor is predicted for the protein. A BLASTP search reveals that the predicted protein encoded by *Tb*927.5.4020 has some similarities to the bacterial surface transporter TolA, although at only 52% identity and 14% query cover, the E-value is poor. The TolA protein is involved in the translocation of colicins, antibiotics produced by *Escherichia coli* to reduce competition from closely related species (Penfold et al., 2012). With no transmembrane domain predicted it is, however, unlikely that

*Tb*927.5.4020 encodes a surface transporter, and may instead be an integral membrane protein.

Table 5.3 Zinc finger proteins identified in the downregulated dataset.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE
Tb927.7.2660	ZC3H20	-1.41	0.0004
Tb927.10.12780	ZC3H37	-0.91	0.0056
Tb927.5.810	ZC3H11	-0.82	0.0154
Tb927.1.2200	Zinc finger protein, putative	-0.62	0.0302
Tb927.10.15610	Zinc finger protein, putative	-0.62	0.0299

Table 5.4 mRNAs bound, and stabilised, by ZC3H20 in the downregulated dataset.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE
Tb927.5.4020	Hypothetical protein	-1.50	0.0005
Tb927.5.2160	Hypothetical protein, conserved	-0.74	0.0057
Tb11.v5.0217	Receptor-type adenylate cyclase (GRESAG 4)	-0.62	0.0037

Both ZC3H37 and ZC3H11 have been demonstrated to bind trypanosome MKT1 (Singh et al., 2014), a cytosolic protein that is associated with polysomes. MKT1 plays a role in stress resistance, but also interacts with a number of other RNA-binding proteins and regulators thus suggesting it could be at the core of a large post-transcriptional regulatory network (Singh et al., 2014). ZC3H11 interacts with MKT1 through its C-terminus, whilst its N-terminus binds, and stabilizes, RNA. It is interesting that ZC3H11 is downregulated in the selected monomorphs compared to the parental pleomorphs, since it is essential for the survival of BSF cells (Droll et al., 2013). ZC3H11 however, was demonstrated to bind chaperones required for the heat shock response during differentiation to the procyclic form and RNAi depletion of the protein resulted in the ~2-fold decrease in abundance of many heat shock protein transcripts, amongst others (Droll et al., 2013; Singh et al., 2014).

Of the 22 transcripts showing a 2-fold decrease in abundance after *ZC3H11* RNAi in BSF cells, only one was found to be downregulated in this dataset, GPEET2. ZC3H37 is likewise associated with increasing translation and/or mRNA stability and this was demonstrated by tethering fragments of the protein to a blasticidin resistance protein marker (Erben et al., 2014). Blasticidin resistance in BSF was increased by an average of 64.7-fold when fragments of ZC3H37 were tethered to the resistance marker.

39 of the 68 significantly downregulated transcripts encoded hypothetical proteins. These 39 hypothetical proteins were analyzed for conserved protein domains or features using the InterPro Protein Sequence Analysis and Classification Tool (Version 71.0, EMBL-EBI) (Table 5.5). *Tb*927.1.5260 was the transcript showing the most downregulation between the selected monomorphs and parental pleomorphs and encoded a hypothetical protein. The protein, however, was predicted to have a GPI-anchor, transmembrane domain, cleavage site and was most highly expressed in proliferative BSF cells, thus making it likely that *Tb*927.1.5260 encodes a VSG or ISG. No one protein domain, or family, was enriched in the downregulated group of hypothetical proteins, however the META domain did occur twice. The META domain (PFAM family, PF03724) is present in a number of small bacterial proteins of unknown function and has been linked to motility, protein secretion and the heat shock response (Ramos et al., 2011). In *Leishmania*, META1 and META2 were shown to be specifically upregulated in the metacyclic promastigote form of the parasite (Ramos et al., 2004; Ramos et al., 2011; Uliana et al., 1999). Whilst overexpression of META1 was associated with increased virulence, META2 was shown to be important in protection against heat shock and stress (Ramos et al., 2011; Uliana et al., 1999). The WD40 domain of *Tb*927.7.6300 has interestingly been associated with the regulation of a variety of cellular pathways, including signaling pathways, the cell cycle and translation (Choudhury et al., 2011; Xu and Min, 2011). The WD40 domain features abundantly throughout eukaryotic genomes and is furthermore one of the top interacting domains (Xu and Min, 2011). Within the WD40-repeat protein family are the highly conserved RACK1 proteins. In *Saccharomyces pombe* cells deficient in the RACK1 homologue, *cpc2*, are delayed in cell cycle progression, and, significantly, show a reduced reaction to environmental

stimuli that would normally provoke G1 arrest (McLeod et al., 2000). In mammalian cells, RACK1 physically associates with numerous signal transducers, such as protein kinase C (PKC) (McLeod et al., 2000; Mochly-Rosen et al., 1991). The *T. brucei* homologue of RACK1, TRACK, is expressed throughout the lifecycle and able to functionally complement *cpc2* null mutants of *S. pombe* (Rothberg et al., 2006). RNAi of TRACK demonstrated that TRACK was a required for the initiation of, and progression through, the final stages of mitosis, cytokinesis (in BSF cells) and cell cleavage (in PCF cells) (Rothberg et al., 2006).

Table 5.5 Signal peptides, transmembrane domains and InterPro protein family/domains predicted for the Hypothetical proteins regulated in the selected monomorphs, according to Tritryp.

GENE CODE	CONSERVED?	LOG2FC	SIGNAL PEPTIDE PREDICTED?	TRANSMEMBRANE DOMAINS PREDICTED?	INTERPRO DOMAIN/ FAMILY?
Tb927.1.5260	No	-1.81	No	2	GPI-anchor
Tb927.5.4020	No	-1.50	Yes	No	-
Tb927.5.2260	Yes	-0.93	No	-	META domain
Tb927.7.6330	No	-0.90	No	4	-
Tb927.10.8730	Yes	-0.90	No	-	ABC1 family
Tb927.10.8740	No	-0.88	Yes	1	-
Tb11.v5.0513	Yes	-0.83	No	-	NLI interacting factor-like phosphatase domain
Tb927.10.15620	No	-0.82	Yes	3	-
Tb927.3.1650	No	-0.81	No	2	-
Tb927.11.16820	Yes	-0.80	Yes	-	-
Tb927.11.2410	Yes	-0.80	No	-	BAR/IMD domain-like
Tb927.11.8340	No	-0.79	Yes	3	-
Tb927.11.810	Yes	-0.79	No	-	TPR-like domain
Tb11.1390	Yes	-0.74	Yes	6	-
Tb927.5.2160	Yes	-0.74	No	-	META domain
Tb927.7.6300	Yes	-0.72	No	-	WD40 domain, G-beta repeat
Tb927.3.1640	No	-0.71	No	3	-
Tb927.11.240	Yes	-0.71	Yes	-	-
Tb927.10.5100	No	-0.71	No	3	-
Tb927.9.2320	Yes	-0.71	No	-	Methyltransferase domain
Tb927.8.4300	No	-0.70	Yes	3	-
Tb927.10.12060	Yes	-0.70	No	-	F-box domain
Tb927.3.5760	Yes	-0.69	Yes	-	-
Tb927.11.6040	Yes	-0.67	No	-	MFS general substrate transporter family
Tb927.7.520	Yes	-0.65	No	-	Cytochrome b5-like Heme/Steroid binding domain
Tb927.4.2410	Yes	-0.64	No	-	Glycosyl hydrolase family 65 central catalytic domain
Tb11.v5.0209	Yes	-0.64	Yes	-	Adenylate and Guanylate cyclase catalytic domain
Tb927.11.2480	No	-0.62	No	2	-
Tb927.11.1830	Yes	-0.61	No	6	-

5.5.1.2 Selected monomorphs upregulate genes associated with the proliferative slender bloodstream form

Far fewer transcripts were upregulated in the selected monomorph population relative to the parental pleomorphic population and moreover these changes were to a much lesser extent than the downregulated genes (Appendix D). Only 11 transcripts that showed a significant increase in abundance after the 10-week passage were selected for further analysis. The highest Log2FC recorded was +0.7315 (a fold-change of roughly 1.66) for *Tb*927.7.3250 (ESAG 6), whilst the 10 remaining genes scored Log2FC values between +0.6997 and +0.5018. Over 60% of the transcripts identified as being upregulated encoded either glycosomal or surface proteins, whilst one transcript each encoded the remaining proteins shown in Figure 5.11.

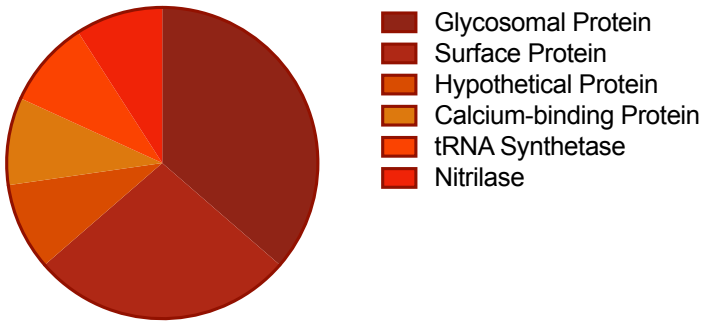


Figure 5.11 Upregulated genes in the selected monomorph population relative to the parental pleomorph population grouped by protein description. Total=11.

Many of the proteins encoded by the transcripts downregulated in the selected monomorphs relative to the parental pleomorphs, such as GPEET and PAD2 are required for correct progression from the proliferative long slender BSF to the transmission-competent stumpy form and onwards to the insect procyclic form. This shift of the transcriptome away from pleomorphism is further corroborated by the upregulation of numerous transcripts encoding proteins primarily associated with the proliferative slender form of the parasite (Table 5.6).

The top two hits in the upregulated dataset are ESAG 6 and ESAG 7, that together form the transferrin receptor (Steverding et al., 1994). Transferrin is

essential for the growth of BSF parasites and must be acquired entirely form the mammalian host. The trypanosome's heterodimeric transferrin receptor (*Tb*TfR) bares no structural resemblance to the human receptor and is instead postulated to have evolved from VSG N-terminal domains (Pays et al., 2001; Salmon et al., 1997).

Table 5.6 Annotated upregulated genes that are associated with the proliferative long slender bloodstream form.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	GENE FUNCTION
Tb927.7.3250	ESAG 6	0.73	0.0277	Transferrin receptor (Steverding et al., 1994).
Tb927.7.3260	ESAG 7	0.70	0.0219	Transferrin receptor (Steverding et al., 1994).
Tb927.9.11580	gim5A	0.59	0.0061	One monomer of the glycosomal membrane protein dimer, GIM5 (Maier et al., 2001).
Tb927.11.5520	Triosephosphate isomerase (TPI or TIM)	0.51	0.0066	Glycosomal enzyme (Swinkels et al., 1986).
Tb927.6.4280	Glyceraldehyde 3-phosphate dehydrogenase	0.50	0.0033	Glycosomal enzyme (Misset et al., 1987).
Tb927.6.4300	Glyceraldehyde 3-phosphate dehydrogenase	0.50	0.0030	Glycosomal enzyme (Misset et al., 1987).

BSF parasites produce ATP by glycolysis, a process that occurs within the membrane-enclosed glycosome (Opperdoes and Borst, 1977). A handful of transcripts encoding glycosomal structural or enzymatic components were identified as being significantly upregulated by the selected monomorphs. Gim5a dimerises together with gim5b to form the constitutively expressed glycosomal membrane protein GIM5, the most abundant protein found within the glycosomal membrane and essential for survival of BSF parasites (Maier et al., 2001), and the glycolytic enzymes identified, triosephosphate isomerase and glyceraldehyde 3-phosphate dehydrogenase, catalyse the interconversion of DHAP and G3P and G3P to BPGA within the glycosome, respectively (Figure 5.12). The *Tb*927.6.4300 GAPDH

transcript is slightly longer than that of *Tb*927.6.4280, otherwise the two transcripts are identical.

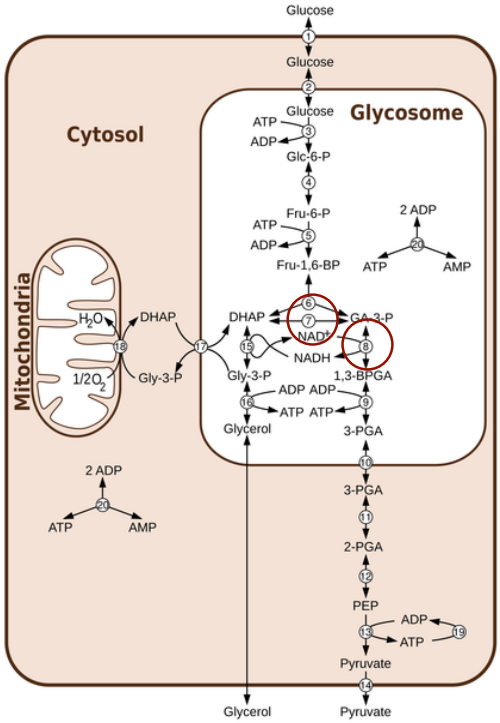


Figure 5.12 Glycolytic enzyme activities of *Tb*927.11.5520 and *Tb*927.6.4280/4300. Aerobic glycolysis within BSF *T. brucei*. *Tb*927.11.5520, triosephosphate isomerase, catalyses step 7 and *Tb*927.6.4280/4300, GAPDH, catalyses step 8. Adapted from (Achcar et al., 2012).

A 44kDa calflagin (Tb44) was identified in the significantly upregulated dataset. Calflagins are a family of calcium-binding proteins containing an EF-hand calcium-binding loop that localise to the flagellar membrane, where they associate with lipid raft microdomains and presumably act as calcium signalling proteins (Emmer et al., 2010; Wu et al., 1992). The expression of Tb44, as well two other calflagins, Tb17 and Tb24, is regulated during the life cycle, with approximately 10-fold greater expression in BSF parasites than PCF (Emmer et al., 2010). The remaining upregulated transcripts encoded an amino acid transporter (*Tb*927.4.4860), the metabolic repair enzyme NIT1 (*Tb*927.9.1960) and asparaginyl-tRNA synthetase (*Tb*927.4.2310).

5.5.1.3 Changes in transcript abundance occur early in selection

In addition to investigating the identity of the genes that showed changes in transcript abundance following prolonged laboratory passage, the RNAseq analysis was furthermore used as a tool to observe how the expression of these genes changed occurred over time. Did the changes in gene expression occur gradually, akin to the acquisition of *in vitro* 8-pCPTcAMP resistance, or were they restricted to a certain period of the passage? The 'Intermediate' samples 1 and 2 (Figure 5.8a) flanked the sudden increase in 8-pCPTcAMP resistance and it was hypothesised that sequencing RNA from these timepoints could reveal the underlying cause for such a rise in 8-pCPTcAMP resistance.

Of the 68 genes that were significantly downregulated in the final selected population relative to the 'Start' population, 55 were also significantly downregulated in the 'Intermediate' 1 population relative to the 'Start' population (Figure 5.13a). Neither the 'Intermediate 2 relative to Intermediate 1' nor the 'End relative to Intermediate 2' comparisons identified significant changes in transcript abundance ($p \leq 0.05$). Of the 13 genes that did not also appear in the 'Intermediate 1 relative to Start' comparison, and therefore only significantly decreased in abundance in the latter stages of selection, 12 encoded hypothetical proteins. The exception to this was *Tb*927.1.2200, which codes for a putative zinc finger protein. The fold change heatmap in Figure 5.13a demonstrates that the vast majority of the changes in transcript abundance occurred in the early stages of selection. Despite most transcriptome changes occurring in the first few weeks of passage, 33 of the 55 genes that appeared in both the 'End relative to Start' and 'Intermediate 1 relative to Start' comparisons showed a slight progression in downregulation after the initial significant drop in transcript abundance. The remaining 22 genes showed stable expression, or slightly decreased expression compared to the 'End' selected population at the 'Intermediate' 1 time point, however in these cases the differences in fold change between the two comparisons were negligible.

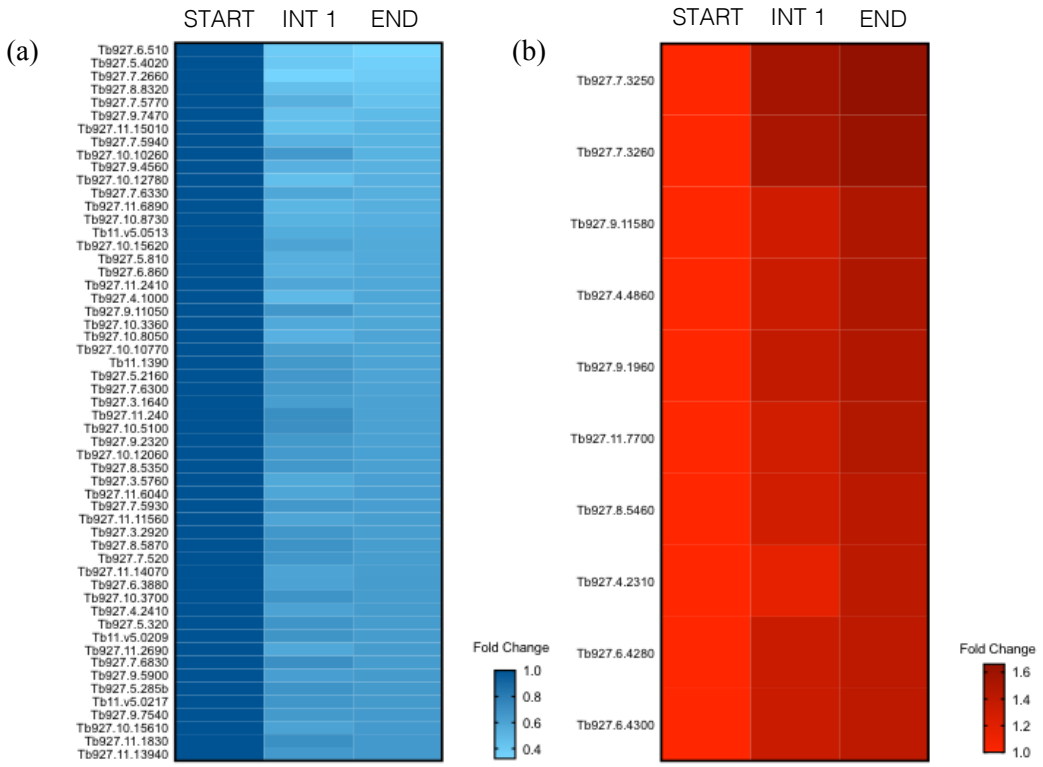


Figure 5.13 Gene expression over three selection timepoints. Fold change of transcript abundance in the 'Intermediate 1' and 'End' populations relative to the starting pleomorphic population, respectively. **(a)** Downregulated genes. The 55 genes that were significantly downregulated in both the 'End relative to Start' and 'Intermediate 1 relative to Start' comparisons are plotted in blue. **(b)** Upregulated genes. The 10 genes that were significantly upregulated in both the 'End relative to Start' and 'Intermediate 1 relative to Start' comparisons are plotted in red. Each row represents a different transcript, denoted by the gene codes on the left. The transcripts are ordered from the highest fold change to the lowest. No significant changes were observed in the 'Intermediate 2 relative to Intermediate 1' and 'End relative to Intermediate 2' comparisons.

The changes in expression of the 11 upregulated transcripts followed a similar pattern (Figure 5.13b). Again, no significant changes in transcript abundance were observed in the 'Intermediate 2 relative to Intermediate 1' or 'End relative to Intermediate 2' comparisons. Ten genes were significantly downregulated in the 'Intermediate 1 to Start' comparison. The transcript encoding triosephosphate isomerase, *Tb*927.11.5520, was the only transcript not identified in the 'Intermediate

1 relative to Start' comparison. Although most of the largest changes to transcript abundance again materialised early during selection, all ten of the transcripts showed a progression in their upregulation throughout the *in vitro* selection that was more profound that that observed for the downregulated genes. This can be most easily visualised for transcripts *Tb*927.9.11580 (gim5a), *Tb*927.4.4860 (amino acid transporter 8) and *Tb*927.11.7700 (hypothetical protein) (Figure 5.13b).

5.5.1.4 The RNAseq data is validated by qRTPCR

Validation of the RNAseq data was performed by qRTPCR of three downregulated genes, *Tb*927.11.15010 (*NEK21*), *Tb*927.5.4020 (Hypothetical protein) and *Tb*927.7.2660 (*ZC3H20*). The transcript levels in one replicate of the RNA acquired from the parental pleomorphs ('Start') were compared to transcript levels in the triplicate selected monomorph RNA samples ('End 1-3') for each of the three genes (Figure 5.14). The validation was performed on the same material that had been sent for RNA sequencing. The mean fold change of transcript abundance when calculated by qRTPCR closely matched that generated from the RNAseq data analysis (*ZC3H20*- 0.317 vs 0.355, *Tb*927.5.4020- 0.27 vs 0.376, *NEK21*- 0.442 vs 0.499), thus validating the datasets.

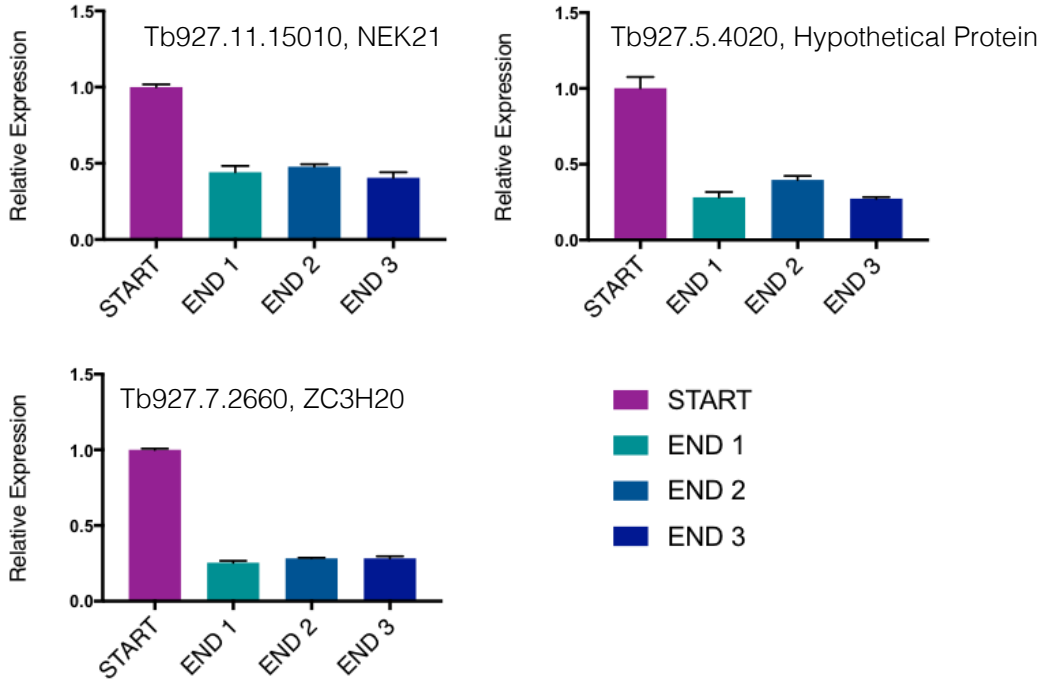


Figure 5.14 Validation of the RNAseq data by qRTPCR. Expression levels of *NEK21*, *Tb927.5.4020* and *ZC3H20* were determined in three RNA replicates from the selected monomorphs ('End 1-3') relative to one of the pleomorphic parental population RNA samples ('Start'). The respective transcripts measured for abundance are inset into each plot. Data represents the mean ± S.D (n=3).

5.5.1.5 Two independent monomorphic *T. brucei* lines differ in gene expression

The monomorphic *T. brucei* Lister 427 line is routinely used in laboratory studies. The life history of these parasites is complicated, but it has been proposed that all current *T. brucei* Lister 427 lines likely originate from the Shinyanga III strain that was isolated from cattle in Tanganyika (present day Tanzania) in 1956 (Cross, 2018). Over 6 decades, the parasite has been extensively passaged under different conditions in different laboratories around the world. As such, it is highly likely that *T. brucei* Lister 427 lines from different laboratories are not phenotypically nor genotypically identical. Since 1956, the parasite has become laboratory adapted and consequently lost its capacity for slender to stumpy to differentiation (Cross, 2018) . In total, the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected

monomorphs significantly downregulated 950 genes relative to the pleomorphic parental population. To study if any of the genes for which expression was selected against in these selected monomorphs had also been selected against in our laboratory's *T. brucei* Lister 427 strain, the primers used in Section 5.4.1.4 were used to quantify *NEK21*, *ZC3H20* and *Tb*927.5.4020 gene expression in a logarithmic *in vitro* population of *T. brucei* Lister 427 cells.

Interestingly, the results in Figure 5.15 show that bloodstream form *T. brucei* Lister 427 cells consistently express greater levels of *NEK21*, *ZC3H20* and *Tb*927.5.4020 than the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs. Whilst NEK21 transcript abundance is only ~1.5-fold greater in the selected monomorph population relative to the *T. brucei* Lister 427 population, both ZC3H20 and *Tb*927.5.4020 are expressed to a much greater extent (~3- and 4.5-fold, respectively). Although the *T. brucei* Lister 427 cell line did express NEK21 and ZC3H20 at a significantly lower level than the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi pleomorphs, the *T. brucei* Lister 427 cells in fact expressed significantly more *Tb*927.5.4020 than the pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population (up 125%). Whilst only a very small fraction of the downregulated genes was tested, these results appear to suggest that the genes identified in this screen as being downregulated upon laboratory adaptation have not also been selected for in another line of monomorphic BSF cells and that in fact other monomorphic lines may express these to a greater extent than pleomorphic cells.

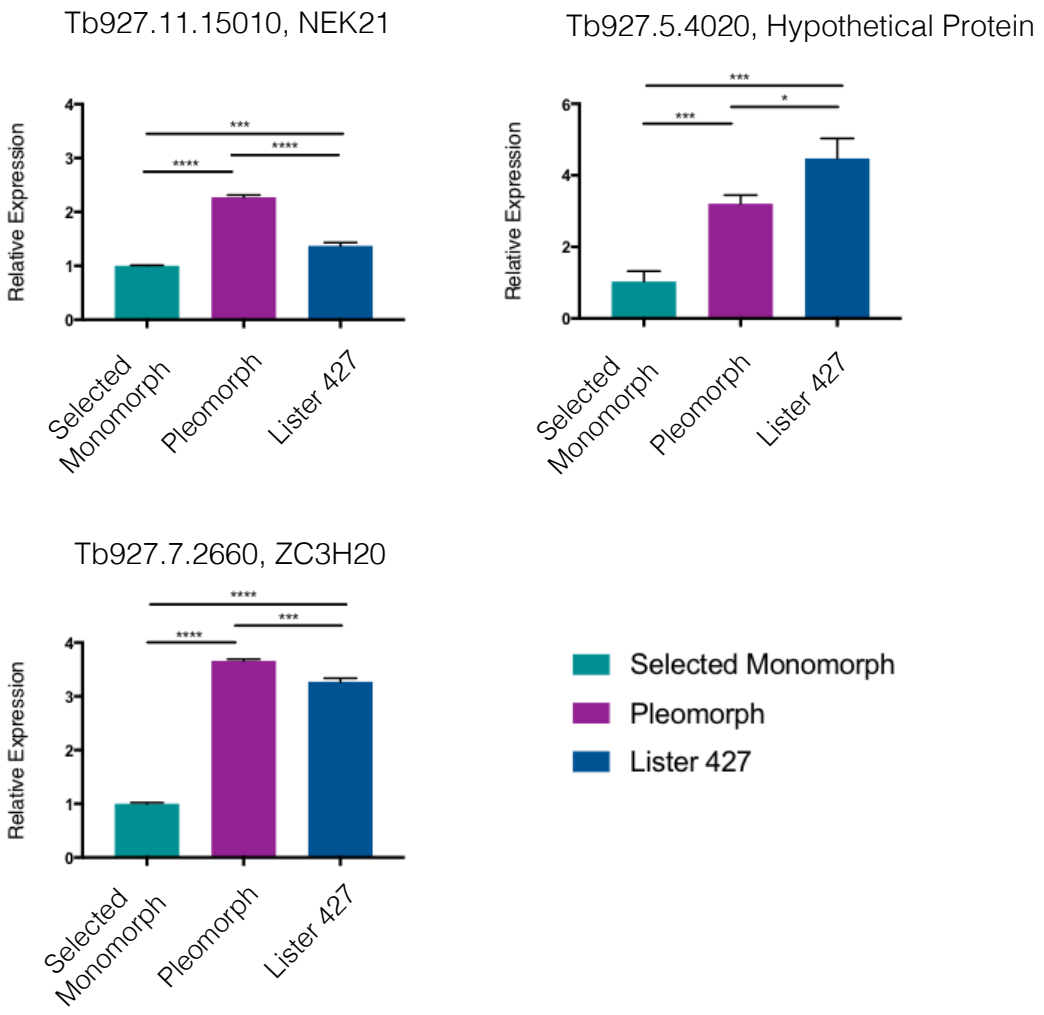


Figure 5.15 *NEK* RNAi GFP^{ESpro}AnTat1.1^{ES} selected monomorphs and *T. brucei* Lister 427 monomorphs do not express similar levels of NEK21, ZC3H20 or *Tb*927.5.4020. Expression was determined by qRTPCR. The respective transcripts measured for abundance are inset into each plot. Data represents the mean ± S.D (n=3).

5.5.2 Genome analysis of selected monomorphs

To try and understand why the observed changes in gene expression may have occurred upon the loss of differentiation competency, DNA from the parental pleomorph 'Start' population and the selected monomorph 'End' population was sent for whole genome sequencing (WGS) and analysed for single nucleotide polymorphisms (SNPs) or insertions/deletions (indels). The processing of raw sequencing data, identification of SNPs/Indels and mapping of SNP regions to the

reference *T. b. brucei* TREU927 genome were kindly performed by Dr Alasdair Ivens (see Section 2.19).

When a PHRED-scaled QUAL score threshold of 200 was applied to the Variant Call Format (VCF) data, a total of 33,609 SNPs and 935 indels were found in the selected monomorph's assembled genome compared to the parental pleomorph's. The 301bp regions containing these SNPs or indels were mapped to the *T. b. brucei* TREU927 reference genome and those which did not fall into the coding sequence (CDS) of a differentially expressed gene identified by the transcriptome analysis filtered from the dataset. Overall, 82 SNPs and 1 indel were identified in the selected monomorph's genome compared to the parental pleomorph's, although the single indel was excluded from further analysis since it occurred in a highly repetitive region and was likely to have been the result of sequencing error. The 82 SNPs occurred over 25 of the genes identified as being significantly downregulated in the selected monomorphs, and 3 of the genes identified as being significantly upregulated in the selected monomorphs compared to the parental pleomorphs (Appendix E). With the exception of the GT to T deletion in *ZC3H37*, for which the selected monomorphs were homozygous for, the selected monomorphs were heterozygous at each SNP site, with one allele encoding the original nucleotide present in the pleomorphic genome and one allele encoding an alternative base (Appendix E).

During the SNP identification process, the reads generated from the WGS of the 'End' selected monomorph population were aligned against what was essentially a haploid 'Start' reference genome which assumed homozygosity at every site. In reality however, trypanosomes are diploid organisms. Therefore, it was important to assess whether there was already underlying genetic variation in the 'Start' population at each of the proposed SNP sites. To achieve this, the reads generated from the WGS of the 'Start' selected monomorph population were aligned against the 'Start' reference genome and the SNPs identified compared to those shown in Appendix E. At each of the previously identified heterozygous SNP sites from the 'End' vs 'Start' analysis, the 'Start' parental pleomorphs were also identified as being heterozygous. Therefore, a SNP within the CDS of one of the differentially

expressed genes identified in the transcriptome analysis did not appear to be responsible for the changes in gene expression between the parental pleomorphs and the selected monomorphs.

The *T. brucei* genome is organised into large polycistronic transcription units (PTUs) (Berriman et al., 2005; Johnson et al., 1987) which are transcribed by RNApol II, with mature mRNA transcripts being processed by coupled *trans*-splicing and polyadenylation (Matthews et al., 1994). The lack of transcriptional control within these PTUs requires that *T. brucei* gene expression is controlled largely at the post-transcriptional level, either through mRNA stability or translational efficiency. Most post-transcriptional control of gene expression is governed through regulatory elements in the 3'UTRs of genes (Berberof et al., 1995; Hotz et al., 1995; Hug et al., 1993), although upstream open reading frames (uORFs) in the 5' UTRs can also influence gene expression (Jensen et al., 2014; Siegel et al., 2005; Vasquez et al., 2014). Therefore, a SNP within either the 5' or 3' UTRs of a gene could significantly impact its expression. To explore whether differential gene expression between the selected monomorphs and parental pleomorphs was the result of a SNP (or SNPs) disrupting a regulatory motif, splice acceptor site or polyadenylation site, the list of SNPs identified in the 'End' vs 'Start' comparison was searched for those which fell within the 5' or 3' UTR of one of the differentially expressed genes. To account for the diploid genome, the SNPs identified in the 'End' vs 'Start' alignment were again compared to those from the 'Start' vs 'Start' alignment, and only those for which there was a difference in genotype between the selected monomorphs and parental pleomorphs analysed further. With these filters in place, only one SNP was identified- a T to A substitution on one allele within the 3'UTR of *Tb*927.8.8320, a hypothetical protein. The genotype of the starting pleomorphic population at this site (TT) however, was called with low confidence. As part of the Variant Call Format output, for each SNP site a 'normalised' PHRED score of the likelihood of each possible genotype is generated. The score of the most likely genotype is 0 and this score increases up to 255 with decreasing likelihood. For this site in the 'Start' genome, a score of 0 was given to the genotype TT, a score of 3 given to the genotype TA and a score of 162 for the genotype AA. Therefore, whilst it was highly unlikely that, at this site, the 'Start' genotype was AA, the

programme could not confidently predict whether this site was TT or TA. This is reflected in the number of reads for either T or A at this site- out of 34 reads, 19 were for T and 12 were for A. In contrast, the genotype at this site in the selected monomorphs' genome was predicted with high confidence (244 for TT, 0 for TA and 209 for AA). Since the reads were split almost equally between T and A at this site, and the genotype confidence score was poor, this SNP was not chosen for further analysis.

Since focusing on the genes differentially expressed between the parental pleomorphs and selected monomorphs had not identified any SNPs which could explain the differences in gene expression, we next turned our attention to the SNP sites which occurred in the CDS of any gene and at which there was a difference in genotype between the pleomorphs and selected monomorphs. If a SNP had occurred in a 'master regulator' gene which did not affect its transcription but had impacted its efficiency, this could explain why the gene was not identified in the transcriptome analysis but gene expression changes were measurable. For this analysis, the PHRED-scaled QUAL score threshold was raised to 999. The Variant Call Format identified 15 SNP sites at which there was a difference in genotype between the parental pleomorphs and the selected monomorphs (Table 5.7). Most of the SNPs occurred within unannotated regions of the genome or pseudogenes. Three SNPs, however, fell within the CDS of a gene and one of these caused a non-synonymous mutation, therefore changing the amino acid sequence of the protein encoded. The H157L mutation in *Tb*927.11.2600, a hypothetical protein which localises to the nucleus (Dean et al., 2017), furthermore changed the properties of the amino acid encoded- from a polar histidine to a non-polar leucine. A search for functional domains and sites using the InterPro Protein Sequence Analysis and Classification Tool (Version 71.0, EMBL-EBI) found that no part of the protein resembled any annotated functional domains, however it did contain 5 predicted regions of disorder across its length (although the SNP did not occur in any of these regions). Expression of *Tb*927.11.2600 did not significantly change between the pleomorphs and selected monomorphs, as determined by the transcriptome analysis. However, as was the case for the SNP site within the UTR of *Tb*927.8.8320, the genotype at the site of the SNP in *Tb*927.11.2600 could not be predicted with high confidence for either the parental

pleomorphs or selected monomorphs. The 'normalised' PHRED scores for the likelihood of each genotype in the parental pleomorphs were 0 for AA (the given genotype), 7 for TA and 255 for TT and for the selected monomorphs the scores were 0 for TA (the given genotype), 13 for AA and 255 for TT. Therefore, whilst it is clear that the genotype at this site is not TT, it is unclear for both populations whether the genotype is AA or TA.

Table 5.7 SNP sites at which there was a difference in genotype between the selected monomorphs and the parental pleomorphs

CONTIG	START GENOTYPE	END GENOTYPE	SNP SITE?	GENE ANNOTATION	SYNONYMOUS OR NON-SYNONYMOUS
709, Pos 693	CC	TC	Non-coding region	-	-
1515, Pos 2908	CA	CC	Region with SNP has BLAST similarity to pseudogenic VSG transcript	-	-
1721, Pos 2041	AG	GG	Non-coding region	-	-
2041, Pos 2526	GG	AG	Non-coding region	-	-
2061, Pos 390	AG	GG	Tb927.2.420	DNA-directed RNA polymerase, putative, pseudogene	-
2103, Pos 3529	CA	CC	Region with SNP has low BLAST similarity to pseudogenic VSG transcript	-	-
2333, Pos 1081	TC	TT	Tb927.1.2880	Pteridine transporter, putative	Synonymous
3664, Pos 2748	GG	AG	Region with SNP has low BLAST similarity to pseudogenic VSG transcript	-	-
3978, Pos 409	AC	CC	Tb927.2.1330	Retrotransposon hot spot protein 6 (RHS6), degenerate	-
4118, Pos 454	GA	AA	Non-coding region	-	-
4143, Pos 1414	GG	AG	Region with SNP has BLAST similarity to pseudogenic VSG transcript	-	-
4808, Pos 2305	AA	TA	Tb927.11.2600	Conserved hypothetical protein, nuclear	Non-synonymous. H157L
4977, Pos 730	CA	AA	Tb927.9.1240	VSG pseudogene	-
5978, Pos 526	CC	TC	Region with SNP has BLAST similarity to Tb10.v4.0188	Hypothetical protein, predicted to be glycosyltransferase	Synonymous
10940, Pos 856	CT	CC	Region with SNP has low BLAST similarity to pseudogenic VSG transcript	-	-

5.6 Discussion

Throughout the literature, it is widely reported that laboratory-adapted monomorphic strains of *T. brucei* switch their expressed VSG at a rate significantly lower than that of differentiation-competent pleomorphic strains. However, it has not been demonstrated whether the loss of pleomorphism and reduction in VSG switch rate associated with laboratory-adaptation occurs concomitantly and therefore are directly linked, or whether the two processes can be selected for independently and thus uncoupled. By using an 'inducible-monomorphism' model, we have demonstrated by *in vitro* FACS-based VSG switch assays and VSGseq that VSG switch frequency and differentiation capacity can be uncoupled. Inducible knockdown of a gene involved in the slender to stumpy quorum sensing pathway (Mony et al., 2014) did not cause a reduction of VSG switch rate in three independent RNAi lines or change the expressed VSG diversity.

In this Chapter, I adopted a second approach to ask whether VSG switch frequency and differentiation capacity can be uncoupled. Populations of selected monomorphs were generated by repeated *in vitro* passage of the pleomorphic reporter cell lines used in Chapters 3 and 4. *In vivo*, the populations of selected monomorphs demonstrated reduced responsiveness to SIF as exemplified by the retention of slender morphology, diminished expression of the stumpy marker protein PAD1 and decreased accumulation in the G0/G1 stage of the cell cycle. When the VSG switch rate of the selected monomorphs was compared to that of the parental pleomorphs using the FACS-based switch assay format it was observed that, although the cells had lost the capacity to differentiate to stumpy forms and therefore become monomorphic, the VSG switch rate had not changed. Therefore, VSG switch rate and slender to stumpy differentiation capacity had been uncoupled. These results corroborate our findings using the inducible monomorphism model. To investigate the changes that had occurred to confer monomorphism during the *in vitro* passage, an 'evolve and re-sequence' approach and RNAseq analysis was adopted. Transcriptome analysis found that the selected monomorphs were depleted of a number of transcripts which were required for the stumpy and procyclic forms of the parasites, such as GPEET and EP1 procyclin, PAD1 and PAD2 and GRESAG 4.

Conversely, a number of transcripts required by the slender bloodstream form stage of the parasite were upregulated, such as the two subunits of the transferrin receptor, ESAGs 6 and 7 and the glycosomal enzymes triosephosphate isomerase and glyceraldehyde 3-phosphate dehydrogenase. In addition, a number of CCCH zinc finger proteins were significantly downregulated in the selected monomorphs compared to the parental pleomorphs. The zinc finger protein ZC3H20 binds to and stabilises its target mRNAs and is already known to bind two developmentally-regulated, PCF-enriched transcripts. ZC3H20 could represent a novel regulator of slender to stumpy differentiation. Whole genome sequencing and SNP/INDEL analysis could not identify any mutations within the differentially expressed genes which would explain the changes in transcription between the parental pleomorphs and selected monomorphs. In Chapter 3, it was observed that the knock down of *HYP2* caused a significant increase in VSG switch rate, from $6.97x10^{-4} \pm 0.0004$ to $2.5x10^{-3} \pm 0.0009$ switches/cell/generation, suggesting that *HYP2* could be involved in the regulation of both antigen switching and differentiation from slender to stumpy forms. These switch assay results were not reproducible, however, and furthermore, VSGseq analysis in Chapter 4 did not present evidence for a greater number of VSGs being expressed in the induced $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2* RNAi population compared to the uninduced population. *HYP2* could therefore be involved in the regulation of monoallelic control (which would only be observed at a single-cell level using the VSGseq approach), as discussed in Section 3.10. Future work should look at the single cell level to establish the role of *HYP2* in antigen switching. The evidence presented in this Chapter, however, clearly demonstrated that despite the loss of efficient differentiation capacity, VSG switch rate in the selected $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2* RNAi population did not significantly change. This result therefore corroborates our results with the $GFP^{ESpro}AnTat1.1^{ES}$ *NEK* and *DYRK* RNAi cell lines to show that antigen switch rate and slender to stumpy differentiation capacity are not mechanistically tethered.

In both this Chapter and in Chapter 3, it was observed that the pleomorphic $GFP^{ESpro}AnTat1.1^{ES}$ *HYP2* and *NEK* RNAi lines consistently switched expression of VSGs at a rate approximately 4 times lower, and more so by *in situ* switches of transcription, than the $GFP^{ESpro}AnTat1.1^{ES}$ *DYRK* RNAi line. Laboratory-adapted

trypanosomes are frequently characterised by low VSG switch rates and infrequent DNA recombination-based VSG switches (Liu et al., 1985; McCulloch et al., 1997; Turner, 1997). It is interesting then, that the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cell line was the first cell line out of the three passaged reporter cell lines to lose the capacity to differentiate to stumpy forms in response to the accumulation of SIF, but without a reduction in VSG switch rate. What this evidence suggests is that, in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cell line, loss of slender to stumpy differentiation capacity and reduced VSG switch rate have been selected for, however at different rates and independently of one another. The data, alongside the *in vitro* 8-pCPTcAMP growth curves in Figure 5.3 where a steady progression in resistance was observed over time, also demonstrates that both reduced VSG switch rate and loss of differentiation capacity are not abrupt 'on-off' events, but rather progress along a spectrum over time (this relies on the assumption that all three reporter RNAi cell lines were generated from the same *T. brucei* AnTat1.1 genetic background and therefore that the VSG switch rate and differentiation capacity at an earlier timepoint were identical). The pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cell line could have demonstrated reduced VSG switch rate compared to, and lost its differentiation capacity sooner than, the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi cell lines because it had been passaged *in vitro* for a longer period of time prior to its use in this study. However, without knowing the culture history of the cell line this is difficult to prove. It would be interesting to continue passage of the selected GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population to observe how long it would take before a significant shift in VSG switch frequency was observed.

The selected monomorphs were depleted of a number of transcripts which were associated with the stumpy and insect stages of the parasite's lifecycle. Rather than an accumulation of mutations in each of these genes being the reason for the loss of pleomorphism, it was more likely that these changes were the result of changes to an upstream regulator (or regulators). Three CCCH zinc finger proteins (ZC3Hs) were identified as being significantly downregulated in the selected monomorphs compared to the parental pleomorphs- ZC3H11, ZC3H20 and ZC3H37 (Table 5.3). Of these, ZC3H20, the fourth most downregulated transcript behind GPEET procyclin and two hypothetical proteins, was the most promising candidate for a role

in the regulation of slender to stumpy differentiation. In addition to its role of stabilising the mRNA of two developmentally-regulated, PCF-enriched transcripts (MCP12 and TS-like E (Ling et al., 2011)), the zinc-finger protein has furthermore been identified in two other unpublished differentiation screens performed in the Matthews Lab at the University of Edinburgh. The first of these screens, performed by Dr Lindsay McDonald, used a genome-wide RNAi library screen to identify downstream effectors of the differentiation-inducing compound GlaxoSmithKline Kinase Inhibitor (GKI) 7 (Diaz et al., 2014). Treatment with 1.44μM GKI7 for 24 hours causes G1 arrest and upregulation of PAD1 protein expression in *T. brucei* AnTat1.1 90:13 cells (Lindsay McDonald Thesis, 2016). In a method similar to the RNAi library screen which identified components of the *T. brucei* quorum sensing pathway (Mony et al., 2014), triplicate induced populations of a monomorphic RNAi library were treated with GKI7 and the RNAi inserts of the GKI7 resistant cells subsequently subjected to Ion Torrent Sequencing. ZC3H20 was identified in one of the three replicates, suggesting that it could be a factor involved in slender to stumpy differentiation. The second screen was performed by Dr Mathieu Cayla and made use of mass-spectrometry to identify differentially phosphorylated proteins in a *T. brucei* AnTat1.1 90:13 Δ*DYRK* cell line. When *DYRK*, a known component of the quorum sensing signalling pathway, was deleted, ZC3H20 was significantly less phosphorylated. Dr Cayla was also able to demonstrate that the protein was phosphorylated when purified *T. brucei* lysate was used as a substrate of a purified active kinase.

The role of ZC3H20 as a regulator of slender to stumpy differentiation was validated *in vivo* by Dr Mathieu Cayla (data unpublished, Figure 5.16). *In vivo*, the parasitaemia of *T. brucei* AnTat1.1 J1339 Δ*ZC3H20* cells ascended rapidly, with one mouse required to be humanely culled on day 5 of the experiment (Figure 5.16a). Despite the high density of parasites in the blood, the *ZC3H20* knock out cells did not show any evidence of accumulation in G0/G1 (Figure 5.16b) or increased expression of PAD1 (Figure 5.16c). This data demonstrates that ZC3H20 plays a role in the regulation of slender to stumpy differentiation and furthermore, validates the transcriptome analysis performed as part of this study. Whether the other ZC3H zinc

finger proteins identified, ZC3H11 and ZC3H37, also contribute to slender to stumpy form differentiation remains to be confirmed *in vivo*.

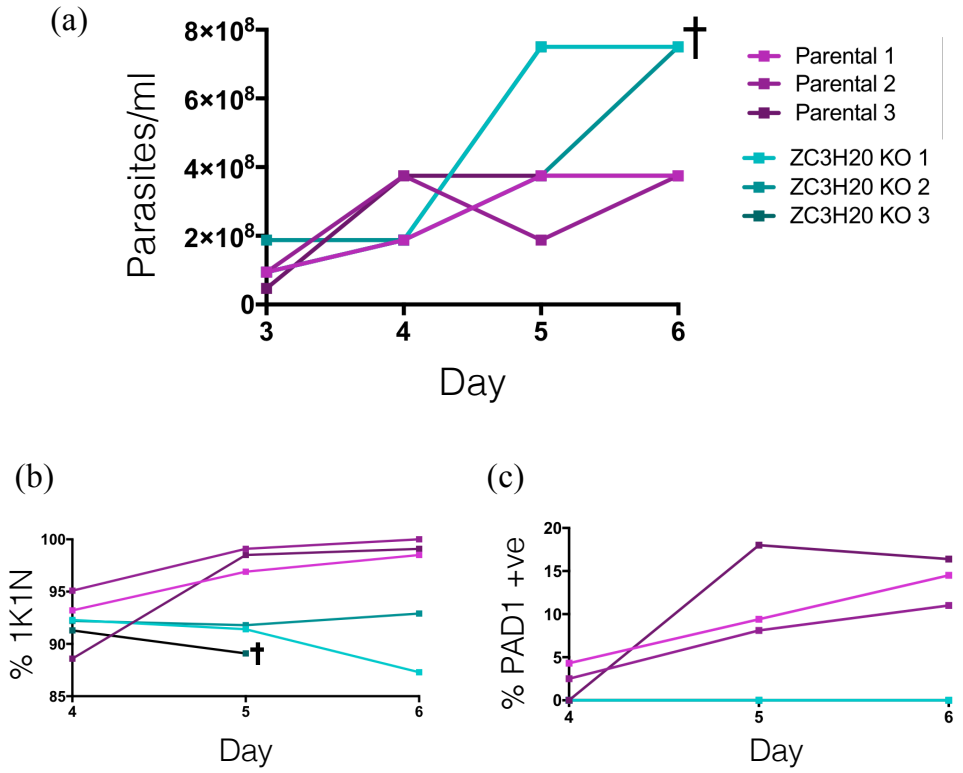


Figure 5.16 ZC3H20 knock out in a pleomorphic cell line confers increased resistance to SIF Generation of the ZC3H20 knock out cell line and collection of all of the above data was performed by Dr Mathieu Cayla (University of Edinburgh). **(a)** *In vivo* growth of *T. brucei* AnTat1.1 J1339 Δ*ZC3H20* cells compared parental cell line. The Δ*ZC3H20* cells remained proliferative, whilst the parental cells arrested in response to SIF. The Δ*ZC3H20* replicate 3 was culled on day 5 of the experiment on humane grounds. **(b)** Cell-cycle analysis of cells in the blood on days 4, 5 and 6 of infection. Pleomorphic cells (purple) arrested in G1/G0 as they differentiated, seen as an accumulation of 1K1N cells between days 4 and 5. The proportion of cells in 1K1N within the Δ*ZC3H20* population (teal) remained relatively constant over the infection time course. At least 80 cells were counted for each individual infection. **(c)** IFA analysis of PAD1 expression. Despite the high parasitaemia at later infection timepoints, no Δ*ZC3H20* cells were found to express the PAD1 protein on their surface.

The Log2FC value of *ZC3H20* in the selected monomorphs compared to the parental pleomorphs was -1.41. Whole genome sequencing and SNP/INDEL analysis however, did not identify any mutations within the genomic sequence of the selected monomorphs' ZC3H20 which could explain the changes in expression. Therefore, it is possible that there may be a further regulator whose expression impacts that of ZC3H20. Analysis of the CDS and UTR sequences of all the differentially expressed genes identified in the transcriptome analysis found only one site, within the 3' UTR of the hypothetical protein, *Tb*927.6.810, where the genotype differed between the parental GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi pleomorphs and the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs. The confidence with which the genotype could be predicted at this site, however, was very poor and therefore there is a strong possibility that the genotype may actually be the same. There was only SNP within the length of the entire selected monomorphs' genome which was predicted to cause a non-synonymous mutation in the CDS of a gene (another hypothetical protein, *Tb*927.11.2600), however in this instance the confidence scores for the genotype predictions in both the parental pleomorph and selected monomorphs populations were poor.

For both the RNAseq and WGS, genomic material prepared from the whole population was sent for sequencing. Although the period of *in vitro* passage began with a clonal pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population, we do not know if, during the ten-week selection period, the population has remained relatively homogenous or whether the selected monomorphs are composed of multiple different genotypes which all confer the same phenotypic traits. That only 15 SNPs which caused a change in genotype at that site were identified across the entire length of the selected monomorphs' genome, suggests that the population could still be homogenous. However, without sequencing the genomes of multiple clones from both populations, this is hard to prove. By performing WGS on clones derived from the selected monomorphs population, we could firstly deduce whether the population is clonal and secondly clarify whether the SNPs in the CDS of *Tb*927.11.2600 and the 3'UTR of *Tb*927.6.810 are real and therefore potentially causing the changes in gene expression.

The SNP/INDEL analysis only focused on those areas of the *T. brucei* TREU 927 genome which are annotated as CDSs or UTRs. Table 5.7 shows that 4 of the 15 SNPs in the selected monomorphs' genome occurred in unannotated regions. It may be, despite their lack of annotation, that one of more of these regions are important in the regulation of slender to stumpy differentiation and their identity remains to be discovered. Alternatively, the differential gene expression between the selected monomorphs and parental pleomorphs could be the result of changes to epigenetic control mechanisms, such as chromatin structure or histone post-translational modifications (methylation, ubiquitylation, phosphorylation and SUMOylation). It is already known that a variety of epigenetic mechanisms are critical to the maintenance of monoallelic VSG expression (see Section 1.2.5) and furthermore, that a histone methyltransferase, DOT1B, is essential for differentiation to the procyclic form of the parasite (Janzen et al., 2006). In addition to the whole genome sequencing of clones derived from the selected monomorph population, future work could ask if the epigenome of the selected monomorphs has been altered relative to the parental pleomorphs.

In addition to the 'End' versus 'Start' comparison, two 'Intermediate' RNA samples, 1 and 2, were sent for sequencing with the aim of observing how any changes in gene expression were occurring over time. The two 'Intermediate' samples were selected for RNAseq analysis since they flanked the timepoint where a sudden jump in the resistance to 8-pCPTcAMP treatment was observed in the passaged GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population (Figure 5.8a). The results of the transcriptome analysis found that the majority of the differentially expressed genes identified in the 'End' versus 'Start' comparison were also significantly up- or down-regulated in the 'Intermediate 1' relative to 'Start' comparison. No genes significantly changed in expression in the 'Intermediate 2' relative to 'Intermediate 1' comparison nor the 'End' relative to 'Intermediate 2' comparisons. From these results, it could be concluded that the changes in gene expression which confer monomorphism happened early in the period of selection. Since only the final population was tested for SIF resistance and we do not know what 'level' of 8-pCPTcAMP resistance corresponds to resistance to SIF, it is not known whether the 'Intermediate 1' sample, which does not differentially express any genes relative to

the ‘End’ selected monomorph population, is also resistant to SIF. Since the transcriptome profiles of the ‘Intermediate 1’ and ‘End’ selected populations are so similar, it seems likely that these cells may have already lost differentiation capacity (providing the genes that we have identified are indeed the ones driving the loss of differentiation capacity). The differentiation capacity of the ‘Intermediate 1’ population should be tested *in vivo*, and if life cycle competency has been lost, the aliquots frozen in the earlier during the passage should also be tested in order to establish exactly where differentiation capacity was lost. What is important to consider, however, is that before RNA could be isolated from the ‘Intermediate 1’ population, the cells had to be thawed and grown to a sufficient density. This freeze/thaw cycle represents a significant bottleneck and also requires approximately a week of *in vitro* culture. In this time frame, the ‘Intermediate 1’ population’s transcriptome could have changed to closer resemble that of the ‘Intermediate 2’ population, thus confusing the analysis. An improvement to any future study performing this kind of *in vitro* selection would involve isolating genomic material at the same time as cell stabilates as frozen.

To ask whether the changes in gene expression which were observed in our selected monomorph line were consistent between different populations of monomorphic trypanosomes, the expression of three genes significantly downregulated in the selected monomorphs (*ZC3H20*, *NEK21* and *Tb*927.5.4020, a binding partner of ZC3H20) was tested by qRTPCR in the well-studied monomorphic *T. brucei* Lister 427 strain. Interestingly, the *T. brucei* Lister 427 cell line expressed each of the three genes to a significantly greater extent than the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs and furthermore expressed *Tb*927.5.4020, a hypothetical protein, at a significantly higher level than the parental GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi pleomorphs (Figure 5.15). In a study by Nilsson et al. (2010), RNAseq was used to analyse transcript abundance in a slender pleomorphic *T. brucei* AnTat1.1 strain and a monomorphic *T. brucei* Lister 427 strain. Just as there were inconsistencies in gene expression between the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs and the *T. brucei* Lister 427 monomorphs for each of the three tested genes, there were also inconsistencies in gene expression (relative to the pleomorphic cells) between the *T. brucei* Lister 427

strain used in this study and the one used by Nilsson et al. (2010). Though Nilsson et al. (2010) also described *ZC3H20* expression to be significantly lower in the *T. brucei* Lister 427 monomorphs compared to the *T. brucei* AnTat1.1 pleomorphs, the results for the remaining two genes are the opposite of what was reported in this study. They found that *NEK21* expression was higher in the *T. brucei* Lister 427 cells compared to *T. brucei* AnTat1.1 cells and furthermore that the *T. brucei* Lister 427 cells expressed less Tb927.5.4020 than pleomorphic *T. brucei* AnTat1.1 cells (Nilsson et al., 2010). These data sets add weight to the conclusions reached in Chapter 3, where it was argued that differences in trypanosome culture techniques between laboratories could induce 'laboratory specific' adaptations and therefore complicate comparisons between different populations of the same genetic background (Section 3.10). Since each of the three monomorphic cell lines expressed the three genes to different extent, this could suggest, as was discussed in Section 3.10, that monomorphism has arisen multiple times by different means, and that 'monomorphism' may actually comprise several genotypes and phenotypes. To test this hypothesis, the whole transcriptomes and genomes of multiple monomorphic *T. brucei* strains would have to be compared. It must also be noted, however, that both the slender *T. brucei* AnTat1.1 and *T. brucei* Lister 427 cells used in the Nilsson et al. (2010) study were derived from mouse blood, not *in vitro* culture, and therefore that this may have impacted gene expression levels. Mulindwa et al. (2018) recently suggested that the largest contributing factor to inconsistencies between measured gene expression were due to differences in RNA preparation (Mulindwa et al., 2018). However, since RNA was prepared in the same way for our $GFP^{ESpro}AnTat1.1^{ES}$ *NEK* RNAi selected monomorphs and *T. brucei* Lister 427 cells, this argues against this hypothesis. Our transcriptome datasets were subject to filtration and the application of defined Log2FC threshold cut-offs prior to analysis (as described in Section 5.5.1). By only considering the transcripts with the highest fold changes and resemblance to already known regulators of differentiation, we have potentially excluded *novel* regulators of differentiation from our dataset. Future work will consider the entire list of differentially regulated transcripts and compare the genes identified in our analysis to those identified in a genome-wide RIT-seq screen as having a loss of fitness phenotype on only differentiation (and not BSF or PCF

growth) (Alsford et al., 2011). Although this approach is limited by the use of a monomorphic RNAi library in the RIT-seq screen (differentiation is measured as slender to PCF form and not slender to stumpy) this approach may still aid in identifying novel regulators which were not considered as part of this analysis. Genes identified in both datasets will be validated as a novel regulator of differentiation by CRISPR-Cas9 knock out in pleomorphic cell lines (Beneke et al., 2017).

Historically, lines of selected monomorphs were generated through rapid syringe passage between rodents every 2-3 days (Turner, 1990) and it was in these lines of trypanosomes that homogenous VSG expression and low switch rates were reported. How might our results have looked like if our selected monomorphs were generated *in vivo* rather than *in vitro*? *In vitro*, it is considerably easier to control the driving selection force-in this case, culture at high density to select for cells that are unresponsive to the density-dependent differentiation signal. In the mouse model, however, the selection pressures would differ in that the density of the parasites in the blood would not be so easy to control nor to push to as high levels without killing the animal. Furthermore, additional selection pressures and bottlenecks, such as the host immune response, migration to the adipose tissue and the efficiency of injection could all influence the outcome of the period of passage. As has been seen in the literature, *in vivo* passage of our reporter cell lines could have generated homogenous, antigenically stable populations which had furthermore lost pleomorphism. In this scenario, however, such would be the additional selection pressures exerting their force on antigen expression that it would be hard to tease apart co-selection of VSG switch rate and slender to stumpy differentiation capacity. Our *in vitro* high density passage strategy is therefore advantageous since it directly tests for mechanistic tethering between VSG switch rate and slender to stumpy differentiation capacity by only exerting selection pressure against pleomorphism. Regarding what the transcriptome of our selected monomorphs might have looked like if they were generated *in vivo* rather than *in vitro*, I do not believe we would have observed identical temporal selection dynamics, nor the exact same lists of significantly up and down-regulated genes being generated. This is because, as discussed previously, the selection pressures in the mouse differ from *in vitro* culture

and monomorphism seems to have been generated by varying means in different trypanosome isolates (Figure 5.15).

In summary, we have generated a population of selected monomorphs through the continuous *in vitro* passage of a pleomorphic *T. brucei* GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi cell line and demonstrated that a reduction in VSG switch rate did not accompany the loss of pleomorphism. Therefore, the results corroborate our observations in Chapters 3 and 4 that slender to stumpy differentiation capacity and VSG switch frequency can be uncoupled. RNAseq analysis identified that the selected monomorphs were depleted of a number of CCCH zinc finger proteins, one of which, ZC3H20, has now been validated *in vivo* as a novel regulator of slender to stumpy differentiation. Further whole genome sequencing of clones derived from the selected monomorph population may shed further light on the changes which occurred to provoke the differential gene expression.

6 : SUMMARY AND FUTURE DIRECTIONS

The cyclical waves of parasitaemia associated with a bloodstream *T. brucei* infection are characterised by antigenic variation (a process by which the parasite switches its expressed VSG surface coat to evade host antibody-mediated clearance) and the density-dependent differentiation of proliferative slender forms to transmission competent, cell-cycle arrested stumpy forms. Laboratory-adapted lines of *T. brucei* have been reported to switch their expressed VSG surface coats at a significantly lower rate than pleomorphic populations of trypanosomes which have recently been transmitted through a tsetse fly (Turner, 1997). Upon laboratory-adaptation, these populations also lose the capacity to differentiate into transmission competent stumpy forms and thus become monomorphic. Until now, it has not been established if the reduction in VSG switch frequency is directly coupled to the loss of slender to stumpy differentiation capacity, or whether the two processes, though co-selected by multiple passage, are in fact independent.

The aim of this thesis was to establish if there is a direct relationship between *T. brucei* stumpy formation capacity and VSG antigen switch frequency. To address this, I exploited pleomorphic *T. brucei* RNAi cell lines that could each inducibly silence a gene involved in stumpy formation in an approach that we termed 'inducible monomorphism'. Using this model, I sought to compare the VSG switch rate and expressed VSG diversity between uninduced pleomorphic populations and induced monomorphic populations.

A quantifiable *in vitro* flow cytometry-based VSG switch assay found that VSG switch rate did not significantly decrease following the induction of monomorphism and nor was the preferred mechanism of VSG switching altered (Figure 3.14). Sequencing of the expressed VSGs in populations of pleomorphic and induced monomorphic cells demonstrated that the induction of monomorphism did not cause a reduction in expressed VSG diversity (Figure 4.7), nor change the expressed VSG subset (Figure 4.9). To corroborate these results, populations of selected monomorphs were generated by the prolonged *in vitro* passage of the pleomorphic *T. brucei* RNAi cell lines (Figure 5.1). These cells, which had varying degrees of reduced responsiveness to SIF *in vivo* (Figures 5.4-5.6), were tested in the VSG switch assay format and their VSG switch rates compared to the parental

pleomorphic populations. Despite having lost the capacity to effectively respond to SIF, the selected monomorphs were found to switch VSG expression at the same rate as the parental pleomorphic populations (Figure 5.7).

Overall, the results presented within this thesis clearly demonstrate that VSG switch frequency and slender to stumpy differentiation capacity are not coupled, and instead are co-selected for during long term passage. So how then do our results relate to previously published studies that state laboratory-adapted monomorphs switch VSG expression at lower rates than pleomorphic populations? Past trypanosome studies have frequently compared parasite lines derived from independent selections or for which the passage history (which can span decades) is unknown. In contrast however, this study has made use of isogenic populations which allows a direct comparison to be made and rules out the potential for differences in laboratory specific adaptations influence, and complicate, VSG switch study outcomes.

Turner (1997) stated that "the rates of antigenic variation are several orders of magnitude lower in syringe-passaged lines, such as those routinely used in the majority of laboratory studies, compared with most recently fly-transmitted lines". Critically, whilst the populations of trypanosomes compared by Turner (1997) were laboratory-adapted and isogenic, the cell lines must have retained pleomorphism in order to maintain tsetse infectivity and transmissibility. Similarly, the *T. brucei* AnTat1.1 90:13 cell line which was used throughout this study is both pleomorphic and adapted for laboratory culture. Our results showed that progression of these pleomorphic populations to monomorphism did not impact the VSG switch rate (Figure 5.7). Taking into account the conclusions from the 1997 Turner study, perhaps laboratory adaptation (without the loss of pleomorphism) has caused a significant drop in VSG switch rate in our cell line, but we could not observe this since the drop had preceded the beginning of this study. Without having been able to track VSG switch rate during the entirety of the cell line's 'life history', it is impossible to tell. If this explanation did hold true, however, this would provide yet more evidence that monomorphism and VSG switch rate are co-selected independently upon laboratory-adaptation.

To make definitive conclusions about the ability to uncouple VSG switch rate and slender to stumpy differentiation capacity, it could be argued that our selected monomorphs should be transmitted through a tsetse fly and the VSG switch rate measured after transmission, in a repeat of the Turner study (1997). There are, however, considerable setbacks with this approach. Firstly, the transcriptome analysis in Section 5.5.1 demonstrated that the selected monomorphs have downregulated a number of transcripts essential for survival and life cycle progression in the tsetse fly. The commonly used *T. brucei* Lister 427 strain which has also lost slender to stumpy differentiation capacity has been demonstrated to differentiate to procyclic in the midgut of the tsetse fly, however progress no further than this (Peacock et al., 2008; Szoor et al., 2006). It therefore seems realistically very unlikely that the selected populations would survive transmission. Secondly, for the interpretation of our VSG switch assay it is essential that the starting clone expresses VSG AnTat1.1 (to which we have an antibody against). If the selected monomorphs were successfully transmitted by the tsetse fly, the subsequent bloodstream forms may first be expressing a metacyclic VSG and would therefore require passage until the population switched to expression of VSG AnTat1.1. During this period of passage, a change in VSG switch rate could already be selected for and thus complicate he interpretation of results. Finally, if it was hypothesised that slender to stumpy differentiation capacity and antigen switch frequency were indeed linked and that the antigen switch frequency would increase upon transmission through the tsetse fly, this hypothesis would suggest that the cell line switched at extremely high rate when first isolated from tsetse flies since we calculated VSG switch rates between $2.88x10^{-4}$ and $2x10^{-3}$ switches/cell/generation for our selected monomorphs.

It is possible that Turner (1997) may also have observed a lower switch frequency in serially passaged cells due to the indirect nature of the assays used in their study. As discussed in Section 1.2.7, VSG switching experiments that rely on *in vivo* infections and the clearance of 'non-VSG switched' cells by immune methods (such as complement-mediated lysis, as used by Turner(1997)) can underestimate the VSG switch rate as a result of the clearance of new VSG variants by off-target effects. This potential for cross-reactivity therefore limits the comparisons of overall

VSG switch rate between different cell lines. In contrast, our assay directly measured VSG switching through the use of a VSG specific antibody and at a higher resolution- flow cytometry is considerably more sensitive than microscope immunofluorescence analysis and permits the analysis of millions of cells in contrast to hundreds). The VSG switch rate of the populations following tsetse fly transmission increased significantly (perhaps due to the upregulation of different VSG genes in the VSG repertoire as has been observed for members of the *cir* multigene family in *P. c. chabaudi* (Spence et al., 2013)). It would have been interesting to have seen the VSG switch rate of the trypanosome populations measured *prior* to the serial passage and then compared to the VSG switch rates of the passaged and tsetse transmitted populations in order to see the VSG switch frequency decrease with increasing passage, however this was not performed.

As a continuation of this study, it would be interesting to maintain the *in vitro* passage of the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs and to monitor if there was a significant decrease in their VSG switch frequency. The approach used to determine the factors accounting for the loss of pleomorphism in the selected monomorph population could then be adopted to investigate the observed drop in VSG switch frequency (e.g. genomic, transcriptomic analysis). As discussed in Section 1.2.5, a compendium of epigenetic control mechanisms regulates VSG switching. Therefore, it is also possible that a change in VSG switch frequency would be brought about by a change in one or more of these epigenetic factors and thus, in addition to WGS and RNAseq, the epigenomes of the pleomorphic and selected monomorphic cells should be compared. It has been demonstrated that there is an inverse relationship between telomere length and VSG switch frequency, whereby trypanosomes with short telomeres switch more frequently than those with longer telomeres and furthermore, that cells with short telomeres also switch more frequently by gene conversion than those with longer telomeres (Hovel-Miner et al., 2012). The pleomorphic GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi population was observed to consistently switch expression of its VSGs at a rate approximately 4 times lower, and preferentially by *in situ* switches of transcription, than the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi lines (albeit not significantly so) (Figure 3.14). It would be interesting to test if there were differences in telomere length between the three

RNAi lines, or if alternative epigenetic factors accounted for the differences in VSG switch rate and mechanism. Monomorphic lines of trypanosomes have lost their capacity to complete the full lifecycle and therefore elements of their biology could be considered 'biologically irrelevant' when compared to pleomorphic trypanosomes which are more akin to field isolates (with regard to differentiation capacity). This study, however, has shown that populations of inducible monomorphs, generated through the knockdown of a quorum-sensing signalling gene, and selected monomorphs, generated by continuous high density *in vitro* passage of the pleomorphic reporter cell lines, do not switch VSG expression at a rate lower than pleomorphs nor significantly change the proportion of VSG switches elicited by DNA recombination. These results therefore suggest that monomorphic lines of trypanosomes that have *recently* lost pleomorphism are valid models for studying VSG switch rate and mechanism. Most monomorphic cell lines in the literature, however, have been demonstrated to show increased antigen stability and a preference for *in situ* switches of transcription compared to the pleomorphic reporter cell lines used in this study. It therefore must still be acknowledged that additional selection pressures during their continued passage have altered VSG switching in ways that differ them from pleomorphic cell lines. Of course, the pleomorphic cell lines used in this study, despite retaining transmission competence, have been subject to considerable *in vitro* and *in vivo* passage during their passage which may have altered their VSG switching dynamics compared to field strains. A useful study for the trypanosome field would involve studying VSG expression dynamics in the native bovine host.

In addition to demonstrating that VSG switch frequency and pleomorphism can be uncoupled, the work presented here has identified a number of potential novel regulators of slender to stumpy differentiation. An evolve and resequence approach coupled with RNA sequencing was applied to understand the loss of pleomorphism in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorph population. Transcriptome analysis found that the selected monomorphs were depleted in a number of transcripts associated with the stumpy and insect life stages of the parasites (Table 5.1) and, most interestingly, that a number of CCCH zinc finger proteins were also significantly downregulated compared to the parental pleomorphs

(Table 5.3). One of these, ZC3H20, has subsequently been confirmed *in vivo* as a novel regulator of slender to stumpy differentiation (Figure 5.16) and thus validates the transcriptome data collected. Our RNAseq dataset is replete with potential regulators of slender to stumpy differentiation, such as: two additional CCCH zinc finger proteins, ZC3H11 and ZC3H37; a serine/threonine protein kinase, NEK21; an RNA binding protein, RBP32; and many more hypothetical proteins that we do not yet know the function of. Future analysis will implement the recently described CRISPR Cas9 approach (Beneke et al., 2017) to create independent knockout lines for the differentially expressed genes of interest and test these for their *in vitro* and *in vivo* responses to 8-pCPTcAMP and SIF, respectively.

Although a change in the encoded amino acid sequence of a differentiation regulator could underlie the observed changes in gene expression between the selected monomorphs and parental pleomorphs, no SNP anywhere within the selected monomorph genome could be predicted with high confidence (Table 5.7). Experimental evolution under controlled conditions, followed by next-generation re-sequencing of the whole genome, has been instrumental in determining the mutations responsible for a plethora of phenotypes in viral, yeast, bacterial and *Drosophila melanogaster* experimental systems (Long et al., 2015). Though there are less examples in the field of parasitology, E&R approaches have been used to demonstrate that *mdr1* duplication is a contributor to artemisinin combination therapy resistance in *Plasmodium chabaudi* (Rodrigues et al., 2010) and that PBAP2-G is essential for commitment to gametocytogenesis in *Plasmodium berghei* (Sinha et al., 2014). These E&R studies both share one feature in common- the final selected populations were cloned by limiting dilution prior to WGS. In contrast, in this study, total population DNA and RNA was isolated from the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi pleomorph and selected monomorph populations (as discussed in Section 5.6). We do not know if the final selected monomorph population was clonal or consisted of multiple genotypes conferring the same phenotype and therefore this confounds our analysis. Whilst the RNAseq data has since been validated *in vivo* (Figure 5.16), the SNP analysis was not able to confidently predict any changes in genome sequence. This could simply be because there were none, and the differential gene expression was the consequence of a change, or changes to an epigenetic control

mechanism, or perhaps because the selected monomorph population consisted of multiple genotypes. Therefore, before any firm conclusions can be drawn from the WGS analysis, I believe that the selected monomorph population should be cloned by limiting dilution and multiple clones sent for resequencing of the genome. By sequencing clones we can, importantly, ease the analysis of the SNP data, and also establish whether monomorphism arose due one or multiple independent mutations within the population. It would be interesting to continue the high density *in vitro* passage of the GFP^{ESpro}AnTat1.1^{ES} *HYP2* and *DYRK* RNAi selected monomorphs until their *in vivo* differentiation phenotype closely matched that of the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs and then ask by WGS and RNAseq if monomorphism had occurred in the same way in all three independent lines, or if there were differences between the populations.

It has long been reported that monomorphic lines of *T. brucei* switch their expressed VSG at a rate significantly lower than that of recently isolated pleomorphic trypanosomes. However, in this thesis I have demonstrated by multiple approaches that VSG switch frequency and slender to stumpy differentiation capacity can be uncoupled, thus showing that monomorphic trypanosomes are capable of switching VSG expression at rates comparable to pleomorphic populations. The interplay between antigen switching and slender to stumpy differentiation is the major parasite intrinsic determinant of bloodstream infection dynamic, chronicity and transmissibility (Gjini et al., 2010; MacGregor et al., 2011). Mathematical modelling has demonstrated that the production of stumpy forms limits the expressed antigen repertoire within the population since stumpy cells are irreversibly committed to G0/G1 cell cycle arrest and cannot produce new variants (MacGregor et al., 2011). When considering only proliferative slender cells, however, our study has demonstrated that there is no direct mechanistic linkage between VSG switch frequency and differentiation competence. By being able to individually select for each trait, this creates the potential for a multivariate normal distribution of VSG switch rate and relative pleomorphism between different trypanosome isolates. This could be advantageous for the parasite in different host settings where higher or lower antigen switch frequencies could be required to maintain infection chronicity, or in the context of coinfection with competing trypanosome species (such as *T.*

congolense) where the ability to proliferate to higher densities without arresting could confer a growth advantage. The ability to uncouple antigen switch rate and slender to stumpy differentiation capacity may also have facilitated the emergence of trypanosomes such as *T. evansi*, a monomorphic parasite which is mechanistically transmitted by biting and sucking insects, but where the potential for chronic infection is maintained through effective antigenic variation.

Finally, this study has also highlighted the considerable impact of experimental design and laboratory-specific cell stocks on calculated VSG switch frequency and recommends that a comparison of the of the relative VSG switch rate within a given experimental set-up is more informative than a comparison to the published absolute VSG switch rates in other studies. This has important ramifications when considering how we report, measure and compare VSG switch frequencies.

Bibliography

Abbas, A.H., Silva Pereira, S., D'Archivio, S., Wickstead, B., Morrison, L.J., Hall, N., Hertz-Fowler, C., Darby, A.C., and Jackson, A.P. (2018). The Structure of a Conserved Telomeric Region Associated with Variant Antigen Loci in the Blood Parasite Trypanosoma congolense. Genome biology and evolution *10*, 2458-2473.

Achcar, F., Kerkhoven, E.J., Bakker, B.M., Barrett, M.P., and Breitling, R. (2012). Dynamic modelling under uncertainty: the case of Trypanosoma brucei energy metabolism. PLoS computational biology *8*, e1002352.

Adeyemi, O.S., Molefe, N.I., Awakan, O.J., Nwonuma, C.O., Alejolowo, O.O., Olaolu, T., Maimako, R.F., Suganuma, K., Han, Y., and Kato, K. (2018). Metal nanoparticles restrict the growth of protozoan parasites. Artificial cells, nanomedicine, and biotechnology, 1-9.

Aitcheson, N., Talbot, S., Shapiro, J., Hughes, K., Adkin, C., Butt, T., Sheader, K., and Rudenko, G. (2005). VSG switching in Trypanosoma brucei: antigenic variation analysed using RNAi in the absence of immune selection. Mol Microbiol *57*, 1608-1622.

Al-Salabi, M.I., Wallace, L.J., Luscher, A., Maser, P., Candlish, D., Rodenko, B., Gould, M.K., Jabeen, I., Ajith, S.N., and de Koning, H.P. (2007). Molecular interactions underlying the unusually high adenosine affinity of a novel Trypanosoma brucei nucleoside transporter. Molecular pharmacology *71*, 921-929.

Alexandre, S., Paindavoine, P., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Tebabi, P., and Pays, E. (1996). Families of adenylate cyclase genes in Trypanosoma brucei. Mol Biochem Parasitol *77*, 173-182.

Allen, C.L., Goulding, D., and Field, M.C. (2003). Clathrin-mediated endocytosis is essential in Trypanosoma brucei. The EMBO journal *22*, 4991-5002.

Alsford, S., Eckert, S., Baker, N., Glover, L., Sanchez-Flores, A., Leung, K.F., Turner, D.J., Field, M.C., Berriman, M., and Horn, D. (2012). High-throughput decoding of antitrypanosomal drug efficacy and resistance. Nature *482*, 232-236.

Alsford, S., and Horn, D. (2012). Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. Nucleic Acids Res *40*, 10150-10160.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel

sequencing of RNA interference targets in the African trypanosome. Genome Res *21*, 915-924.

Amiguet-Vercher, A., Pérez-Morga, D., Pays, A., Poelvoorde, P., Van Xong, H., Tebabi, P., Vanhamme, L., and Pays, E. (2004). Loss of the mono-allelic control of the VSG expression sites during the development of Trypanosoma brucei in the bloodstream. Molecular Microbiology *51*, 1577-1588.

Aranda, S., Laguna, A., and de la Luna, S. (2011). DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. FASEB J *25*, 449-462.

Aresta-Branco, F., Pimenta, S., and Figueiredo, L.M. (2016). A transcription-independent epigenetic mechanism is associated with antigenic switching in Trypanosoma brucei. Nucleic Acids Res *44*, 3131-3146.

Ashcroft, M.T. (1960). A comparison between a syringe-passaged and a tsetse-fly-transmitted line of a strain of Trypanosoma rhodesiense. Annals of tropical medicine and parasitology *54*, 44-53.

Babokhov, P., Sanyaolu, A.O., Oyibo, W.A., Fagbenro-Beyioku, A.F., and Iriemenam, N.C. (2013). A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. Pathogens and global health *107*, 242-252.

Barnwell, E.M., van Deursen, F.J., Jeacock, L., Smith, K.A., Maizels, R.M., Acosta-Serrano, A., and Matthews, K. (2010). Developmental regulation and extracellular release of a VSG expression-site-associated gene product from Trypanosoma brucei bloodstream forms. J Cell Sci *123*, 3401-3411.

Barquilla, A., Saldivia, M., Diaz, R., Bart, J.M., Vidal, I., Calvo, E., Hall, M.N., and Navarro, M. (2012). Third target of rapamycin complex negatively regulates development of quiescence in Trypanosoma brucei. Proc Natl Acad Sci U S A *109*, 14399-14404.

Barrett, M.P., Boykin, D.W., Brun, R., and Tidwell, R.R. (2007). Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. British journal of pharmacology *152*, 1155-1171.

Barry, J.D., Crowe, J.S., and Vickerman, K. (1983). Instability of the Trypanosoma brucei rhodesiense metacyclic variable antigen repertoire. Nature *306*, 699-701.

Barry, J.D., and Emery, D.L. (1984). Parasite development and host responses during the establishment of *Trypanosoma brucei* infection transmitted by tsetse fly. Parasitology *88*, 67-84.

Barry, J.D., Marcello, L., Morrison, L.J., Read, A.F., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U., Caler, E.*, et al.* (2005). What the genome sequence is revealing about trypanosome antigenic variation. Biochemical Society transactions *33*, 986-989.

Bartossek, T., Jones, N.G., Schafer, C., Cvitkovic, M., Glogger, M., Mott, H.R., Kuper, J., Brennich, M., Carrington, M., Smith, A.S.*, et al.* (2017). Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. Nature microbiology *2*, 1523-1532.

Bassler, B.L. (2002). Small talk. Cell-to-cell communication in bacteria. Cell *109*, 421-424.

Batram, C., Jones, N.G., Janzen, C.J., Markert, S.M., and Engstler, M. (2014). Expression site attenuation mechanistically links antigenic variation and development in Trypanosoma brucei. Elife *3*, e02324.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017). A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. Royal Society open science *4*, 170095.

Benz, C., Mulindwa, J., Ouna, B., and Clayton, C. (2011). The Trypanosoma brucei zinc finger protein ZC3H18 is involved in differentiation. Mol Biochem Parasitol *177*, 148-151.

Berberof, M., Perez-Morga, D., and Pays, E. (2001). A receptor-like flagellar pocket glycoprotein specific to Trypanosoma brucei gambiense. Mol Biochem Parasitol *113*, 127-138.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S., and Pays, E. (1995). The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. The EMBO journal *14*, 2925-2934.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B.*, et al.* (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416-422.

Berriman, M., Hall, N., Sheader, K., Bringaud, F., Tiwari, B., Isobe, T., Bowman, S., Corton, C., Clark, L., Cross, G.A.*, et al.* (2002). The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. Mol Biochem Parasitol *122*, 131-140.

Berthier, D., Breniere, S.F., Bras-Goncalves, R., Lemesre, J.L., Jamonneau, V., Solano, P., Lejon, V., Thevenon, S., and Bucheton, B. (2016). Tolerance to Trypanosomatids: A Threat, or a Key for Disease Elimination? Trends Parasitol *32*, 157-168.

Besteiro, S., Barrett, M.P., Riviere, L., and Bringaud, F. (2005). Energy generation in insect stages of Trypanosoma brucei: metabolism in flux. Trends Parasitol *21*, 185-191.

Bitter, W., Gerrits, H., Kieft, R., and Borst, P. (1998). The role of transferrin-receptor variation in the host range of Trypanosoma brucei. Nature *391*, 499-502.

Black, S.J., Guirnalda, P., Frenkel, D., Haynes, C., and Bockstal, V. (2010). Induction and regulation of Trypanosoma brucei VSG-specific antibody responses. Parasitology *137*, 2041-2049.

Bochud-Allemann, N., and Schneider, A. (2002). Mitochondrial substrate level phosphorylation is essential for growth of procyclic Trypanosoma brucei. J Biol Chem *277*, 32849-32854.

Boothroyd, C.E., Dreesen, O., Leonova, T., Ly, K.I., Figueiredo, L.M., Cross, G.A., and Papavasiliou, F.N. (2009). A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. Nature *459*, 278-281.

Breidbach, T., Ngazoa, E., and Steverding, D. (2002). Trypanosoma brucei: in vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. Experimental parasitology *101*, 223-230.

Briggs, E., Crouch, K., Lemgruber, L., Lapsley, C., and McCulloch, R. (2018). Ribonuclease H1-targeted R-loops in surface antigen gene expression sites can direct trypanosome immune evasion. PLoS Genet *14*, e1007729.

Brown, R.C., Evans, D.A., and Vickerman, K. (1972). Developmental changes in ultrastructure and physiology of Trypanosoma brucei. Transactions of the Royal Society of Tropical Medicine and Hygiene *66*, 336-337.

Brown, R.C., Evans, D.A., and Vickerman, K. (1973). Changes in oxidative metabolism and ultrastructure accompanying differentiation of the mitochondrion in Trypanosoma brucei. Int J Parasitol *3*, 691-704.

Bruce, D., Harvey, D., Hamerton, A.E., Davey, J.B., and Bruce, L. (1912). The morphology of the trypanosome causing disease in man in Nyasaland. Proc R Soc Lond B *85*, 423-433.

Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014). Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. J Biol Chem *289*, 20273-20282.

Bulow, R., Overath, P., and Davoust, J. (1988). Rapid lateral diffusion of the variant surface glycoprotein in the coat of Trypanosoma brucei. Biochemistry *27*, 2384-2388.

Bursell, E. (1963). Aspects of the metabolism of amino acids in the tsetse fly, *Glossina* (Diptera). J Ins Physiol *9*, 439-452.

Buscher, P., Bart, J.M., Boelaert, M., Bucheton, B., Cecchi, G., Chitnis, N., Courtin, D., Figueiredo, L.M., Franco, J.R., Grebaut, P.*, et al.* (2018). Do Cryptic Reservoirs Threaten Gambiense-Sleeping Sickness Elimination? Trends Parasitol *34*, 197-207.

Caljon, G., Van Reet, N., De Trez, C., Vermeersch, M., Perez-Morga, D., and Van Den Abbeele, J. (2016). The Dermis as a Delivery Site of Trypanosoma brucei for Tsetse Flies. PLoS Pathog *12*, e1005744.

Campbell, D.A., van Bree, M.P., and Boothroyd, J.C. (1984). The 5'-limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking (TAA)90. Nucleic Acids Res *12*, 2759-2774.

Capewell, P., Clucas, C., DeJesus, E., Kieft, R., Hajduk, S., Veitch, N., Steketee, P.C., Cooper, A., Weir, W., and MacLeod, A. (2013a). The TgsGP gene is essential for resistance to human serum in Trypanosoma brucei gambiense. PLoS Pathog *9*, e1003686.

Capewell, P., Cren-Travaille, C., Marchesi, F., Johnston, P., Clucas, C., Benson, R.A., Gorman, T.A., Calvo-Alvarez, E., Crouzols, A., Jouvion, G.*, et al.* (2016). The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. Elife *5*.

Capewell, P., Monk, S., Ivens, A., Macgregor, P., Fenn, K., Walrad, P., Bringaud, F., Smith, T.K., and Matthews, K.R. (2013b). Regulation of Trypanosoma brucei Total and Polysomal mRNA during Development within Its Mammalian Host. PloS one *8*, e67069.

Capewell, P., Veitch, N.J., Turner, C.M., Raper, J., Berriman, M., Hajduk, S.L., and MacLeod, A. (2011). Differences between Trypanosoma brucei gambiense groups 1 and 2 in their resistance to killing by trypanolytic factor 1. PLoS Negl Trop Dis *5*, e1287.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., and Turner, M. (1991). Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. Journal of molecular biology *221*, 823-835.

Cavalcanti, D.P., and de Souza, W. (2018). The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy. Scanning *2018*, 9603051.

Cestari, I., and Stuart, K. (2018). Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. Current genomics *19*, 119-132.

Chattopadhyay, A., Jones, N.G., Nietlispach, D., Nielsen, P.R., Voorheis, H.P., Mott, H.R., and Carrington, M. (2005). Structure of the C-terminal domain from Trypanosoma brucei variant surface glycoprotein MITat1.2. J Biol Chem *280*, 7228-7235.

Checchi, F., Filipe, J.A., Haydon, D.T., Chandramohan, D., and Chappuis, F. (2008). Estimates of the duration of the early and late stage of gambiense sleeping sickness. BMC infectious diseases *8*, 16.

Choudhury, K., Cardenas, D., Pullikuth, A.K., Catling, A.D., Aiyar, A., and Kelly, B.L. (2011). Trypanosomatid RACK1 orthologs show functional differences associated with translation despite similar roles in Leishmania pathogenesis. PloS one *6*, e20710.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. Protein engineering *12*, 1035-1040.

Cross, G.A. (1996). Antigenic variation in trypanosomes: secrets surface slowly. BioEssays : news and reviews in molecular, cellular and developmental biology *18*, 283-291.

Cross, G.A. (2018). Pedigrees of some Trypanosoma brucei strains in common use and an index of Lister 427 VSGs.

Cross, G.A., Kim, H.S., and Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Mol Biochem Parasitol *195*, 59-73.

Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I., and Borst, P. (2002). J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. Mol Microbiol *46*, 37-47.

Cross, M., Taylor, M.C., and Borst, P. (1998). Frequent loss of the active site during variant surface glycoprotein expression site switching in vitro in Trypanosoma brucei. Molecular and cellular biology *18*, 198-205.

Czichos, J., Nonnengaesser, C., and Overath, P. (1986). Trypanosoma brucei: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. Experimental parasitology *62*, 283-291.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T.*, et al.* (2011). The variant call format and VCFtools. Bioinformatics (Oxford, England) *27*, 2156-2158.

Davies, K.P., Carruthers, V.B., and Cross, G.A. (1997). Manipulation of the vsg co-transposed region increases expression-site switching in Trypanosoma brucei. Mol Biochem Parasitol *86*, 163-177.

De Greef, C., and Hamers, R. (1994). The serum resistance-associated (SRA) gene of Trypanosoma brucei rhodesiense encodes a variant surface glycoprotein-like protein. Mol Biochem Parasitol *68*, 277-284.

De Pablos, L.M., Kelly, S., de Freitas Nascimento, J., Sunter, J., and Carrington, M. (2017). Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding proteins in Trypanosoma brucei. Open biology *7*.

Dean, S., Marchetti, R., Kirk, K., and Matthews, K.R. (2009). A surface transporter family conveys the trypanosome differentiation signal. Nature *459*, 213-217.

Dean, S., Sunter, J.D., and Wheeler, R.J. (2017). TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. Trends Parasitol *33*, 80-82.

Deeks, E.D. (2019). Fexinidazole: First Global Approval. Drugs *79*, 215-220.

DeJesus, E., Kieft, R., Albright, B., Stephens, N.A., and Hajduk, S.L. (2013). A single amino acid substitution in the group 1 Trypanosoma brucei gambiense haptoglobin-hemoglobin receptor abolishes TLF-1 binding. PLoS Pathog *9*, e1003317.

Denninger, V., Fullbrook, A., Bessat, M., Ersfeld, K., and Rudenko, G. (2010). The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in Trypanosoma brucei. Mol Microbiol *78*, 459-474.

Denninger, V., and Rudenko, G. (2014). FACT plays a major role in histone dynamics affecting VSG expression site control in Trypanosoma brucei. Mol Microbiol *94*, 945-962.

Devlin, R., Marques, C.A., Paape, D., Prorocic, M., Zurita-Leal, A.C., Campbell, S.J., Lapsley, C., Dickens, N., and McCulloch, R. (2016). Mapping replication dynamics in Trypanosoma brucei reveals a link with telomere transcription and antigenic variation. Elife *5*.

Diaz, R., Luengo-Arratta, S.A., Seixas, J.D., Amata, E., Devine, W., Cordon-Obras, C., Rojas-Barros, D.I., Jimenez, E., Ortega, F., Crouch, S.*, et al.* (2014). Identification and characterization of hundreds of potent and selective inhibitors of Trypanosoma brucei growth from a kinase-targeted library screening campaign. PLoS Negl Trop Dis *8*, e3253.

DiPaolo, C., Kieft, R., Cross, M., and Sabatini, R. (2005). Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Molecular cell *17*, 441-451.

DND*i* (2019a). Acoziborole (Drugs for Neglected Diseases *initiative*).

DND*i* (2019b). Fexinidazole (HAT) (Drugs for Neglected Diseases *initiative*).

Domenicali Pfister, D., Burkard, G., Morand, S., Renggli, C.K., Roditi, I., and Vassella, E. (2006). A Mitogen-activated protein kinase controls differentiation of bloodstream forms of Trypanosoma brucei. Eukaryot Cell *5*, 1126-1135.

Doyle, J.J., Hirumi, H., Hirumi, K., Lupton, E.N., and Cross, G.A. (1980). Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. Parasitology *80*, 359-369.

Drain, J., Bishop, J.R., and Hajduk, S.L. (2001). Haptoglobin-related protein mediates trypanosome lytic factor binding to trypanosomes. J Biol Chem *276*, 30254-30260.

Dreesen, O., and Cross, G.A. (2008). Telomere length in Trypanosoma brucei. Experimental parasitology *118*, 103-110.

Dreesen, O., Li, B., and Cross, G.A. (2007). Telomere structure and function in trypanosomes: a proposal. Nat Rev Microbiol *5*, 70-75.

Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R., and Clayton, C. (2013). Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. PLoS Pathog *9*, e1003286.

DuBois, K.N., Alsford, S., Holden, J.M., Buisson, J., Swiderski, M., Bart, J.M., Ratushny, A.V., Wan, Y., Bastin, P., Barry, J.D.*, et al.* (2012). NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. PLoS Biol *10*, e1001287.

Dyer, N.A., Rose, C., Ejeh, N.O., and Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. Trends Parasitol *29*, 188-196.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., and Donelson, J.E. (2000). The African trypanosome genome. Int J Parasitol *30*, 329-345.

Emmer, B.T., Daniels, M.D., Taylor, J.M., Epting, C.L., and Engman, D.M. (2010). Calflagin inhibition prolongs host survival and suppresses parasitemia in Trypanosoma brucei infection. Eukaryot Cell *9*, 934-942.

Engstler, M., and Boshart, M. (2004). Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. Genes & Development *18*, 2798-2811.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., and Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell *131*, 505-515.

Eperon, G., Balasegaram, M., Potet, J., Mowbray, C., Valverde, O., and Chappuis, F. (2014). Treatment options for second-stage gambiense human African trypanosomiasis. Expert review of anti-infective therapy *12*, 1407-1417.

Erben, E.D., Fadda, A., Lueong, S., Hoheisel, J.D., and Clayton, C. (2014). A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. PLoS Pathog *10*, e1004178.

Eyford, B.A., Ahmad, R., Enyaru, J.C., Carr, S.A., and Pearson, T.W. (2013). Identification of Trypanosome proteins in plasma from African sleeping sickness patients infected with T. b. rhodesiense. PloS one *8*, e71463.

Fairlamb, A.H., and Horn, D. (2018). Melarsoprol Resistance in African Trypanosomiasis. Trends Parasitol *34*, 481-492.

Field, M.C., and Carrington, M. (2009). The trypanosome flagellar pocket. Nat Rev Microbiol *7*, 775-786.

Figueiredo, L.M., and Cross, G.A. (2010). Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryot Cell *9*, 148-154.

Figueiredo, L.M., Janzen, C.J., and Cross, G.A. (2008). A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biol *6*, e161.

Franco, J.R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Mattioli, R.C., and Argaw, D. (2017). Monitoring the elimination of human African trypanosomiasis: Update to 2014. PLoS Negl Trop Dis *11*, e0005585.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014a). Epidemiology of human African trypanosomiasis. Clinical epidemiology *6*, 257-275.

Franco, J.R., Simarro, P.P., Diarra, A., Ruiz-Postigo, J.A., and Jannin, J.G. (2014b). The journey towards elimination of gambiense human African trypanosomiasis: not far, nor easy. Parasitology *141*, 748-760.

Garrett, S., Menold, M.M., and Broach, J.R. (1991). The Saccharomyces cerevisiae YAK1 gene encodes a protein kinase that is induced by arrest early in the cell cycle. Molecular and cellular biology *11*, 4045-4052.

Geigy, R., Huber, M., Weinman, D., and Wyatt, G.R. (1959). Demonstration of trehalose in the vector of African trypanosomiasis: the tsetse fly. Acta tropica *16*, 255-262.

Gerrits, H., Mussmann, R., Bitter, W., Kieft, R., and Borst, P. (2002). The physiological significance of transferrin receptor variations in Trypanosoma brucei. Mol Biochem Parasitol *119*, 237-247.

Giordani, F., Morrison, L.J., Rowan, T.G., HP, D.E.K., and Barrett, M.P. (2016). The animal trypanosomiases and their chemotherapy: a review. Parasitology *143*, 1862-1889.

Gjini, E., Haydon, D.T., Barry, J.D., and Cobbold, C.A. (2010). Critical Interplay between Parasite Differentiation, Host Immunity, and Antigenic Variation in Trypanosome Infections. The American Naturalist *176*, 424-439.

Glover, L., Alsford, S., Beattie, C., and Horn, D. (2007). Deletion of a trypanosome telomere leads to loss of silencing and progressive loss of terminal DNA in the absence of cell cycle arrest. Nucleic Acids Res *35*, 872-880.

Glover, L., Alsford, S., and Horn, D. (2013a). DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. PLoS Pathog *9*, e1003260.

Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. Proc Natl Acad Sci U S A.

Glover, L., Hutchinson, S., Alsford, S., McCulloch, R., Field, M.C., and Horn, D. (2013b). Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. Cell Microbiol *15*, 1984-1993.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology *29*, 644-652.

Gray, A.R. (1965). Antigenic variation in a strain of Trypanosoma brucei transmitted by Glossina morsitans and G. palpalis. Journal of general microbiology *41*, 195-214.

Grunfelder, C.G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.D., Boshart, M., and Overath, P. (2002). Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. Traffic (Copenhagen, Denmark) *3*, 547-559.

Grunfelder, C.G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.D., Morgan, G.W., Field, M.C., and Overath, P. (2003). Endocytosis of a glycosylphosphatidylinositol-anchored protein via clathrin-coated vesicles, sorting by default in endosomes, and exocytosis via RAB11-positive carriers. Molecular biology of the cell *14*, 2029-2040.

Gull, K. (1999). The cytoskeleton of trypanosomatid parasites. Annual review of microbiology *53*, 629-655.

Haanstra, J.R., van Tuijl, A., van Dam, J., van Winden, W., Tielens, A.G., van Hellemond, J.J., and Bakker, B.M. (2012). Proliferating bloodstream-form Trypanosoma brucei use a negligible part of consumed glucose for anabolic processes. Int J Parasitol *42*, 667-673.

Hall, J.P., Wang, H., and Barry, J.D. (2013). Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. PLoS Pathog *9*, e1003502.

Hamm, B., Schindler, A., Mecke, D., and Duszenko, M. (1990). Differentiation of Trypanosoma brucei bloodstream trypomastigotes from long slender to short stumpy-like forms in axenic culture. Mol Biochem Parasitol *40*, 13-22.

Hammarton, T.C., Clark, J., Douglas, F., Boshart, M., and Mottram, J.C. (2003). Stage-specific differences in cell cycle control in Trypanosoma brucei revealed by RNA interference of a mitotic cyclin. J Biol Chem *278*, 22877-22886.

Hammarton, T.C., Monnerat, S., and Mottram, J.C. (2007). Cytokinesis in trypanosomatids. Curr Opin Microbiol *10*, 520-527.

Hartley, C.L., and McCulloch, R. (2008). Trypanosoma brucei BRCA2 acts in antigenic variation and has undergone a recent expansion in BRC repeat number that is important during homologous recombination. Mol Microbiol *68*, 1237-1251.

Hedley, B.D., and Keeney, M. (2013). Technical issues: flow cytometry and rare event analysis. International journal of laboratory hematology *35*, 344-350.

Henard, C.A., Tapscott, T., Crawford, M.A., Husain, M., Doulias, P.T., Porwollik, S., Liu, L., McClelland, M., Ischiropoulos, H., and Vazquez-Torres, A. (2014). The 4-

cysteine zinc-finger motif of the RNA polymerase regulator DksA serves as a thiol switch for sensing oxidative and nitrosative stress. Mol Microbiol *91*, 790-804.

Hendriks, E.F., and Matthews, K.R. (2005). Disruption of the developmental programme of Trypanosoma brucei by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Mol Microbiol *57*, 706-716.

Hendriks, E.F., Robinson, D.R., Hinkins, M., and Matthews, K.R. (2001). A novel CCCH protein which modulates differentiation of Trypanosoma brucei to its procyclic form. The EMBO journal *20*, 6700-6711.

Herbert, W.J., and Lumsden, W.H. (1976). Trypanosoma brucei: a rapid "matching" method for estimating the host's parasitemia. Experimental parasitology *40*, 427-431.

Hertz-Fowler, C., Figueiredo, L.M., Quail, M.A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I.*, et al.* (2008). Telomeric expression sites are highly conserved in Trypanosoma brucei. PloS one *3*, e3527.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. The Journal of parasitology *75*, 985-989.

Horn, D. (2014). Antigenic variation in African trypanosomes. Mol Biochem Parasitol *195*, 123-129.

Horn, D., and Barry, J.D. (2005). The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. Chromosome Res *13*, 525-533.

Horn, D., and Cross, G.A. (1995). A developmentally regulated position effect at a telomeric locus in Trypanosoma brucei. Cell *83*, 555-561.

Horn, D., and Cross, G.A. (1997). Analysis of Trypanosoma brucei vsg expression site switching in vitro. Mol Biochem Parasitol *84*, 189-201.

Hotz, H.R., Lorenz, P., Fischer, R., Krieger, S., and Clayton, C. (1995). Role of 3'-untranslated regions in the regulation of hexose transporter mRNAs in Trypanosoma brucei. Mol Biochem Parasitol *75*, 1-14.

Hovel-Miner, G., Mugnier, M.R., Goldwater, B., Cross, G.A., and Papavasiliou, F.N. (2016). A Conserved DNA Repeat Promotes Selection of a Diverse Repertoire of Trypanosoma brucei Surface Antigens from the Genomic Archive. PLoS Genet *12*, e1005994.

Hovel-Miner, G.A., Boothroyd, C.E., Mugnier, M., Dreesen, O., Cross, G.A., and Papavasiliou, F.N. (2012). Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. PLoS Pathog *8*, e1002900.

Hsia, R., Beals, T., and Boothroyd, J.C. (1996). Use of chimeric recombinant polypeptides to analyse conformational, surface epitopes on trypanosome variant surface glycoproteins. Mol Microbiol *19*, 53-63.

Huang, J., and van der Ploeg, L.H. (1991). Maturation of polycistronic pre-mRNA in Trypanosoma brucei: analysis of trans splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. Molecular and cellular biology *11*, 3180-3190.

Hug, M., Carruthers, V.B., Hartmann, C., Sherman, D.S., Cross, G.A., and Clayton, C. (1993). A possible role for the 3'-untranslated region in developmental regulation in Trypanosoma brucei. Mol Biochem Parasitol *61*, 87-95.

Ilboudo, H., Bras-Goncalves, R., Camara, M., Flori, L., Camara, O., Sakande, H., Leno, M., Petitdidier, E., Jamonneau, V., and Bucheton, B. (2014). Unravelling human trypanotolerance: IL8 is associated with infection control whereas IL10 and TNFalpha are associated with subsequent disease development. PLoS Pathog *10*, e1004469.

Imhof, S., Knusel, S., Gunasekera, K., Vu, X.L., and Roditi, I. (2014). Social motility of African trypanosomes is a property of a distinct life-cycle stage that occurs early in tsetse fly transmission. PLoS Pathog *10*, e1004493.

Jackson, A.P., Allison, H.C., Barry, J.D., Field, M.C., Hertz-Fowler, C., and Berriman, M. (2013). A cell-surface phylome for African trypanosomes. PLoS Negl Trop Dis *7*, e2121.

Jamonneau, V., Ilboudo, H., Kabore, J., Kaba, D., Koffi, M., Solano, P., Garcia, A., Courtin, D., Laveissiere, C., Lingue, K.*, et al.* (2012). Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. PLoS Negl Trop Dis *6*, e1691.

Janzen, C.J., Hake, S.B., Lowell, J.E., and Cross, G.A. (2006). Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in Trypanosoma brucei. Molecular cell *23*, 497-507.

Jensen, B.C., Ramasamy, G., Vasconcelos, E.J., Ingolia, N.T., Myler, P.J., and Parsons, M. (2014). Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. BMC genomics *15*, 911.

Jensen, B.C., Sivam, D., Kifer, C.T., Myler, P.J., and Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC genomics *10*, 482.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273-281.

Jones, N.G., Thomas, E.B., Brown, E., Dickens, N.J., Hammarton, T.C., and Mottram, J.C. (2014). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. PLoS Pathog *10*, e1003886.

Jude, F., Kohler, T., Branny, P., Perron, K., Mayer, M.P., Comte, R., and van Delden, C. (2003). Posttranscriptional Control of Quorum-Sensing-Dependent Virulence Genes by DksA in Pseudomonas aeruginosa. Journal of Bacteriology *185*, 3558-3566.

Kabani, S., Fenn, K., Ross, A., Ivens, A., Smith, T.K., Ghazal, P., and Matthews, K. (2009). Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. BMC genomics *10*, 427.

Kamau, S.W., Grimm, F., and Hehl, A.B. (2001). Expression of green fluorescent protein as a marker for effects of antileishmanial compounds in vitro. Antimicrobial agents and chemotherapy *45*, 3654-3656.

Kassem, A., Pays, E., and Vanhamme, L. (2014). Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in Trypanosoma brucei. Proc Natl Acad Sci U S A *111*, 8943-8948.

Kennedy, P.G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). The Lancet Neurology *12*, 186-194.

Kieft, R., Capewell, P., Turner, C.M., Veitch, N.J., MacLeod, A., and Hajduk, S. (2010). Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. Proc Natl Acad Sci U S A *107*, 16137-16141.

Kim, H.S., and Cross, G.A. (2010). TOPO3alpha influences antigenic variation by monitoring expression-site-associated VSG switching in Trypanosoma brucei. PLoS Pathog *6*, e1000992.

Koomey, M. (1997). Bacterial pathogenesis: a variation on variation in Lyme disease. Curr Biol *7*, R538-540.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics (Oxford, England) *23*, 1289-1291.

La Greca, F., and Magez, S. (2011). Vaccination against trypanosomiasis: can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? Human vaccines *7*, 1225-1233.

LaCount, D.J., Gruszynski, A.E., Grandgenett, P.M., Bangs, J.D., and Donelson, J.E. (2003). Expression and function of the Trypanosoma brucei major surface protease (GP63) genes. J Biol Chem *278*, 24658-24664.

Lamont, G.S., Tucker, R.S., and Cross, G.A. (1986). Analysis of antigen switching rates in Trypanosoma brucei. Parasitology *92 (Pt 2)*, 355-367.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Langousis, G., and Hill, K.L. (2014). Motility and more: the flagellum of Trypanosoma brucei. Nat Rev Microbiol *12*, 505-518.

Laxman, S., Riechers, A., Sadilek, M., Schwede, F., and Beavo, J.A. (2006). Hydrolysis products of cAMP analogs cause transformation of Trypanosoma brucei from slender to stumpy-like forms. Proc Natl Acad Sci U S A *103*, 19194-19199.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) *25*, 2078-2079.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.C., and Kain, S.R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. J Biol Chem *273*, 34970-34975.

Ling, A.S., Trotter, J.R., and Hendriks, E.F. (2011). A zinc finger protein, TbZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes. J Biol Chem *286*, 20152-20162.

Liu, A.Y., Michels, P.A., Bernards, A., and Borst, P. (1985). Trypanosome variant surface glycoprotein genes expressed early in infection. Journal of molecular biology *182*, 383-396.

Liu, D., Albergante, L., Newman, T.J., and Horn, D. (2018). Faster growth with shorter antigens can explain a VSG hierarchy during African trypanosome infections: a feint attack by parasites. Sci Rep *8*, 10922.

Long, A., Liti, G., Luptak, A., and Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature reviews Genetics *16*, 567-582.

Lopez, M.A., Saada, E.A., and Hill, K.L. (2015). Insect stage-specific adenylate cyclases regulate social motility in African trypanosomes. Eukaryot Cell *14*, 104-112.

Lopez-Farfan, D., Bart, J.M., Rojas-Barros, D.I., and Navarro, M. (2014). SUMOylation by the E3 ligase TbSIZ1/PIAS1 positively regulates VSG expression in Trypanosoma brucei. PLoS Pathog *10*, e1004545.

Lowell, J.E., Kaiser, F., Janzen, C.J., and Cross, G.A. (2005). Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei. J Cell Sci *118*, 5721-5730.

Lukes, J., Guilbride, D.L., Votypka, J., Zikova, A., Benne, R., and Englund, P.T. (2002). Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell *1*, 495-502.

Luria, S.E., and Delbruck, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics *28*, 491-511.

Lythgoe, K.A., Morrison, L.J., Read, A.F., and Barry, J.D. (2007). Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. Proc Natl Acad Sci U S A *104*, 8095-8100.

MacGregor, P., and Matthews, K.R. (2012). Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in Trypanosoma brucei. Nucleic Acids Res *40*, 7705-7717.

MacGregor, P., Savill, N.J., Hall, D., and Matthews, K.R. (2011). Transmission stages dominate trypanosome within-host dynamics during chronic infections. Cell Host Microbe *9*, 310-318.

MacGregor, P., Szoor, B., Savill, N.J., and Matthews, K.R. (2012). Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. Nat Rev Microbiol *10*, 431-438.

Magez, S., Caljon, G., Tran, T., Stijlemans, B., and Radwanska, M. (2010). Current status of vaccination against African trypanosomiasis. Parasitology *137*, 2017-2027.

Magez, S., Schwegmann, A., Atkinson, R., Claes, F., Drennan, M., De Baetselier, P., and Brombacher, F. (2008). The role of B-cells and IgM antibodies in parasitemia, anemia, and VSG switching in Trypanosoma brucei-infected mice. PLoS Pathog *4*, e1000122.

Maier, A., Lorenz, P., Voncken, F., and Clayton, C. (2001). An essential dimeric membrane protein of trypanosome glycosomes. Mol Microbiol *39*, 1443-1451.

Maishman, L., Obado, S.O., Alsford, S., Bart, J.M., Chen, W.M., Ratushny, A.V., Navarro, M., Horn, D., Aitchison, J.D., Chait, B.T.*, et al.* (2016). Co-dependence between trypanosome nuclear lamina components in nuclear stability and control of gene expression. Nucleic Acids Res *44*, 10554-10570.

Mancini, P.E., and Patton, C.L. (1981). Cyclic 3',5'-adenosine monophosphate levels during the developmental cycle of Trypanosoma brucei brucei in the rat. Mol Biochem Parasitol *3*, 19-31.

Mantilla, B.S., Marchese, L., Casas-Sanchez, A., Dyer, N.A., Ejeh, N., Biran, M., Bringaud, F., Lehane, M.J., Acosta-Serrano, A., and Silber, A.M. (2017). Proline Metabolism is Essential for Trypanosoma brucei brucei Survival in the Tsetse Vector. PLoS Pathog *13*, e1006158.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res *17*, 1344-1352.

Matthews, K.R. (2005). The developmental cell biology of Trypanosoma brucei. J Cell Sci *118*, 283-290.

Matthews, K.R., McCulloch, R., and Morrison, L.J. (2015). The within-host dynamics of African trypanosome infections. Philos Trans R Soc Lond B Biol Sci *370*.

Matthews, K.R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev *8*, 491-501.

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. Nucleic Acids Res *34*, 5312-5324.

Mbang-Benet, D.E., Sterkers, Y., Crobu, L., Sarrazin, A., Bastien, P., and Pages, M. (2015). RNA interference screen reveals a high proportion of mitochondrial proteins essential for correct cell cycle progress in Trypanosoma brucei. BMC genomics *16*, 297.

McCulloch, R., and Barry, J.D. (1999). A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. Genes Dev *13*, 2875-2888.

McCulloch, R., Cobbold, C.A., Figueiredo, L.M., Jackson, A., Morrison, L.J., Mugnier, M.R., Papavasiliou, F.N., Schnaufer, A., and Matthews, K.R. (2017). Emerging challenges in understanding trypanosome antigenic variation. Emerging Topics in Life Sciences *1*, 585-592.

McCulloch, R., and Field, M.C. (2015). Quantitative sequencing confirms VSG diversity as central to immune evasion by Trypanosoma brucei. Trends in Parasitology *31*, 346-349.

McCulloch, R., Morrison, L.J., and Hall, J.P. (2015). DNA Recombination Strategies During Antigenic Variation in the African Trypanosome. Microbiol Spectr *3*, Mdna3-0016-2014.

McCulloch, R., Rudenko, G., and Borst, P. (1997). Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. Molecular and cellular biology *17*, 833-843.

McDonald, L., Cayla, M., Ivens, A., Mony, B.M., MacGregor, P., Silvester, E., McWilliam, K., and Matthews, K.R. (2018). Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes. PLoS Pathog *14*, e1007145.

McLeod, M., Shor, B., Caporaso, A., Wang, W., Chen, H., and Hu, L. (2000). Cpc2, a fission yeast homologue of mammalian RACK1 protein, interacts with Ran1 (Pat1) kinase To regulate cell cycle progression and meiotic development. Molecular and cellular biology *20*, 4016-4027.

McLintock, L.M., Turner, C.M., and Vickerman, K. (1993). Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy morphology. Parasite immunology *15*, 475-480.

Mesu, V., Kalonji, W.M., Bardonneau, C., Mordt, O.V., Blesson, S., Simon, F., Delhomme, S., Bernhard, S., Kuziena, W., Lubaki, J.F.*, et al.* (2018). Oral fexinidazole for late-stage African Trypanosoma brucei gambiense trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial. Lancet *391*, 144-154.

Michelotti, E.F., and Hajduk, S.L. (1987). Developmental regulation of trypanosome mitochondrial gene expression. J Biol Chem *262*, 927-932.

Miller, E.N., and Turner, M.J. (1981). Analysis of antigenic types appearing in first relapse populations of clones of Trypanosoma brucei. Parasitology *82*, 63-80.

Mochly-Rosen, D., Khaner, H., and Lopez, J. (1991). Identification of intracellular receptor proteins for activated protein kinase C. Proc Natl Acad Sci U S A *88*, 3997-4000.

Mony, B.M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D., and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. Nature *505*, 681-685.

Mony, B.M., and Matthews, K.R. (2015). Assembling the components of the quorum sensing pathway in African trypanosomes. Mol Microbiol *96*, 220-232.

Moreira-Leite, F.F., Sherwin, T., Kohl, L., and Gull, K. (2001). A trypanosome structure involved in transmitting cytoplasmic information during cell division. Science *294*, 610-612.

Morris, J.C., Wang, Z., Drew, M.E., and Englund, P.T. (2002). Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. The EMBO journal *21*, 4429-4438.

Morrison, L.J., Majiwa, P., Read, A.F., and Barry, J.D. (2005). Probabilistic order in antigenic variation of Trypanosoma brucei. Int J Parasitol *35*, 961-972.

Morrison, L.J., Marcello, L., and McCulloch, R. (2009). Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell Microbiol *11*, 1724-1734.

Mugnier, M.R., Cross, G.A., and Papavasiliou, F.N. (2015). The in vivo dynamics of antigenic variation in Trypanosoma brucei. Science *347*, 1470-1473.

Mugnier, M.R., Stebbins, C.E., and Papavasiliou, F.N. (2016). Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma brucei. PLoS Pathog *12*, e1005784.

Mugo, E., and Clayton, C. (2017). Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. PLoS Pathog *13*, e1006560.

Mulindwa, J., Leiss, K., Ibberson, D., Kamanyi Marucha, K., Helbig, C., Melo do Nascimento, L., Silvester, E., Matthews, K., Matovu, E., Enyaru, J.*, et al.* (2018). Transcriptomes of Trypanosoma brucei rhodesiense from sleeping sickness patients, rodents and culture: Effects of strain, growth conditions and RNA preparation methods. PLoS Negl Trop Dis *12*, e0006280.

Muller, L.S.M., Cosentino, R.O., Forstner, K.U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C.J., Arampatzi, P., Vogel, J., Steinbiss, S.*, et al.* (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature.

Myler, P.J., Allen, A.L., Agabian, N., and Stuart, K. (1985). Antigenic variation in clones of Trypanosoma brucei grown in immune-deficient mice. Infection and immunity *47*, 684-690.

Navarro, M., and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. Nature *414*, 759-763.

Nguyen, T.N., Muller, L.S., Park, S.H., Siegel, T.N., and Gunzl, A. (2014). Promoter occupancy of the basal class I transcription factor A differs strongly between active and silent VSG expression sites in Trypanosoma brucei. Nucleic Acids Res *42*, 3164-3176.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., and Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. PLoS Pathog *6*, e1001037.

Nolan, D.P., Rolin, S., Rodriguez, J.R., Van Den Abbeele, J., and Pays, E. (2000). Slender and stumpy bloodstream forms of Trypanosoma brucei display a differential response to extracellular acidic and proteolytic stress. European journal of biochemistry *267*, 18-27.

Odiit, M., Kansiime, F., and Enyaru, J.C. (1997). Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. East African medical journal *74*, 792-795.

Ogbadoyi, E., Ersfeld, K., Robinson, D., Sherwin, T., and Gull, K. (2000). Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. Chromosoma *108*, 501-513.

Opperdoes, F.R., and Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: the glycosome. FEBS letters *80*, 360-364.

Ouna, B.A., Stewart, M., Helbig, C., and Clayton, C. (2012). The Trypanosoma brucei CCCH zinc finger proteins ZC3H12 and ZC3H13. Mol Biochem Parasitol *183*, 184-188.

Paterou, A., Walrad, P., Craddy, P., Fenn, K., and Matthews, K. (2006). Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. J Biol Chem *281*, 39002-39013.

Pays, E., Delronche, M., Lheureux, M., Vervoort, T., Bloch, J., Gannon, F., and Steinert, M. (1980). Cloning and characterization of DNA sequences complementary to messenger ribonucleic acids coding for the synthesis of two surface antigens of Trypanosoma brucei. Nucleic Acids Res *8*, 5965-5981.

Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N., and Steinert, M. (1985). Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. Nature *316*, 562-564.

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Perez-Morga, D. (2001). The VSG expression sites of Trypanosoma brucei: multipurpose tools for the adaptation of the parasite to mammalian hosts. Mol Biochem Parasitol *114*, 1-16.

Pays, E., Van Assel, S., Laurent, M., Darville, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983). Gene conversion as a mechanism for antigenic variation in trypanosomes. Cell *34*, 371-381.

Peacock, L., Ferris, V., Bailey, M., and Gibson, W. (2008). Fly transmission and mating of Trypanosoma brucei brucei strain 427. Mol Biochem Parasitol *160*, 100-106.

Penfold, C.N., Li, C., Zhang, Y., Vankemmelbeke, M., and James, R. (2012). Colicin A binds to a novel binding site of TolA in the Escherichia coli periplasm. Biochemical Society transactions *40*, 1469-1474.

Perez-Morga, D., Vanhollebeke, B., Paturiaux-Hanocq, F., Nolan, D.P., Lins, L., Homble, F., Vanhamme, L., Tebabi, P., Pays, A., Poelvoorde, P.*, et al.* (2005). Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. Science *309*, 469-472.

Peters, J., Fowler, E., Gatton, M., Chen, N., Saul, A., and Cheng, Q. (2002). High diversity and rapid changeover of expressed var genes during the acute phase of Plasmodium falciparum infections in human volunteers. Proc Natl Acad Sci U S A *99*, 10689-10694.

Pinger, J., Chowdhury, S., and Papavasiliou, F.N. (2017). Variant surface glycoprotein density defines an immune evasion threshold for African trypanosomes undergoing antigenic variation. Nature communications *8*, 828.

Pollastri, M.P. (2018). Fexinidazole: A New Drug for African Sleeping Sickness on the Horizon. Trends Parasitol *34*, 178-179.

Portman, N., and Gull, K. (2012). Proteomics and the Trypanosoma brucei cytoskeleton: advances and opportunities. Parasitology *139*, 1168-1177.

Proudfoot, C., and McCulloch, R. (2005). Distinct roles for two RAD51-related genes in Trypanosoma brucei antigenic variation. Nucleic Acids Res *33*, 6906-6919.

Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J.D., and Clayton, C. (2009). Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC genomics *10*, 495.

Radwanska, M., Guirnalda, P., De Trez, C., Ryffel, B., Black, S., and Magez, S. (2008). Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. PLoS Pathog *4*, e1000078.

Ramos, C.S., Franco, F.A., Smith, D.F., and Uliana, S.R. (2004). Characterisation of a new Leishmania META gene and genomic analysis of the META cluster. FEMS Microbiol Lett *238*, 213-219.

Ramos, C.S., Yokoyama-Yasunaka, J.K., Guerra-Giraldez, C., Price, H.P., Mortara, R.A., Smith, D.F., and Uliana, S.R. (2011). Leishmania amazonensis META2 protein confers protection against heat shock and oxidative stress. Experimental parasitology *127*, 228-237.

Redmond, S., Vadivelu, J., and Field, M.C. (2003). RNAit: an automated web-based tool for the selection of RNAi targets in Trypanosoma brucei. Mol Biochem Parasitol *128*, 115-118.

Reuner, B., Vassella, E., Yutzy, B., and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Mol Biochem Parasitol *90*, 269-280.

Rico, E., Rojas, F., Mony, B.M., Szoor, B., Macgregor, P., and Matthews, K.R. (2013). Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei. Frontiers in cellular and infection microbiology *3*, 78.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Roberts, D.J., Craig, A.G., Berendt, A.R., Pinches, R., Nash, G., Marsh, K., and Newbold, C.I. (1992). Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature *357*, 689-692.

Robinson, D.R., Sherwin, T., Ploubidou, A., Byard, E.H., and Gull, K. (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation,

and cytokinesis in the trypanosome cell cycle. The Journal of cell biology *128*, 1163-1172.

Robinson, N.P., Burman, N., Melville, S.E., and Barry, J.D. (1999). Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Molecular and cellular biology *19*, 5839-5846.

Robinson, N.P., McCulloch, R., Conway, C., Browitt, A., and Barry, J.D. (2002). Inactivation of Mre11 does not affect VSG gene duplication mediated by homologous recombination in Trypanosoma brucei. J Biol Chem *277*, 26185-26193.

Roditi, I., Schwarz, H., Pearson, T.W., Beecroft, R.P., Liu, M.K., Richardson, J.P., Buhring, H.J., Pleiss, J., Bulow, R., Williams, R.O*., et al.* (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei. The Journal of cell biology *108*, 737-746.

Rodrigues, L.A., Henriques, G., Borges, S.T., Hunt, P., Sanchez, C.P., Martinelli, A., and Cravo, P. (2010). Experimental evolution of resistance to artemisinin combination therapy results in amplification of the mdr1 gene in a rodent malaria parasite. PloS one *5*, e11593.

Rolin, S., Hancocq-Quertier, J., Paturiaux-Hanocq, F., Nolan, D.P., and Pays, E. (1998). Mild acid stress as a differentiation trigger in Trypanosoma brucei. Mol Biochem Parasitol *93*, 251-262.

Rolin, S., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Nolan, D., Salmon, D., Webb, H., Carrington, M., Voorheis, P., and Pays, E. (1996). Simultaneous but independent activation of adenylate cyclase and glycosylphosphatidylinositol-phospholipase C under stress conditions in Trypanosoma brucei. J Biol Chem *271*, 10844-10852.

Rothberg, K.G., Burdette, D.L., Pfannstiel, J., Jetton, N., Singh, R., and Ruben, L. (2006). The RACK1 homologue from Trypanosoma brucei is required for the onset and progression of cytokinesis. J Biol Chem *281*, 9781-9790.

Rudenko, G. (2000). The polymorphic telomeres of the African Trypanosome trypanosoma brucei. Biochemical Society transactions *28*, 536-540.

Saada, E.A., Kabututu, Z.P., Lopez, M., Shimogawa, M.M., Langousis, G., Oberholzer, M., Riestra, A., Jonsson, Z.O., Wohlschlegel, J.A., and Hill, K.L. (2014). Insect stage-specific receptor adenylate cyclases are localized to distinct subdomains of the Trypanosoma brucei Flagellar membrane. Eukaryot Cell *13*, 1064-1076.

Salmon, D., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Pays, A., Tebabi, P., Nolan, D.P., Michel, A., and Pays, E. (1997). Characterization of the ligand-binding site of the transferrin receptor in Trypanosoma brucei demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. The EMBO journal *16*, 7272-7278.

Salmon, D., Paturiaux-Hanocq, F., Poelvoorde, P., Vanhamme, L., and Pays, E. (2005). Trypanosoma brucei: growth differences in different mammalian sera are not due to the species-specificity of transferrin. Experimental parasitology *109*, 188-194.

Salmon, D., Vanwalleghem, G., Morias, Y., Denoeud, J., Krumbholz, C., Lhomme, F., Bachmaier, S., Kador, M., Gossmann, J., Dias, F.B.*, et al.* (2012). Adenylate cyclases of Trypanosoma brucei inhibit the innate immune response of the host. Science *337*, 463-466.

Sanchez, M.A., Drutman, S., van Ampting, M., Matthews, K., and Landfear, S.M. (2004). A novel purine nucleoside transporter whose expression is up-regulated in the short stumpy form of the Trypanosoma brucei life cycle. Mol Biochem Parasitol *136*, 265-272.

Schulz, D., Mugnier, M.R., Boothroyd, C.E., and Papavasiliou, F.N. (2016). Detection of Trypanosoma brucei Variant Surface Glycoprotein Switching by Magnetic Activated Cell Sorting and Flow Cytometry. Journal of visualized experiments : JoVE.

Schwede, A., Jones, N., Engstler, M., and Carrington, M. (2011). The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Mol Biochem Parasitol *175*, 201-204.

Schwede, A., Macleod, O.J., MacGregor, P., and Carrington, M. (2015). How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins? PLoS Pathog *11*, e1005259.

Seifert, H.S., Wright, C.J., Jerse, A.E., Cohen, M.S., and Cannon, J.G. (1994). Multiple gonococcal pilin antigenic variants are produced during experimental human infections. The Journal of clinical investigation *93*, 2744-2749.

Serkin, C.D., and Seifert, H.S. (2000). Iron availability regulates DNA recombination in Neisseria gonorrhoeae. Mol Microbiol *37*, 1075-1086.

Shapiro, S.Z., Naessens, J., Liesegang, B., Moloo, S.K., and Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. Acta tropica *41*, 313-323.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K., and Rudenko, G. (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. Proc Natl Acad Sci U S A *102*, 8716-8721.

Siegel, T.N., Tan, K.S., and Cross, G.A. (2005). Systematic study of sequence motifs for RNA trans splicing in Trypanosoma brucei. Molecular and cellular biology *25*, 9586-9594.

Silvester, E., McWilliam, K.R., and Matthews, K.R. (2017). The Cytological Events and Molecular Control of Life Cycle Development of Trypanosoma brucei in the Mammalian Bloodstream. Pathogens (Basel, Switzerland) *6*.

Singh, A., Minia, I., Droll, D., Fadda, A., Clayton, C., and Erben, E. (2014). Trypanosome MKT1 and the RNA-binding protein ZC3H11: interactions and potential roles in post-transcriptional regulatory networks. Nucleic Acids Res *42*, 4652-4668.

Sinha, A., Hughes, K.R., Modrzynska, K.K., Otto, T.D., Pfander, C., Dickens, N.J., Religa, A.A., Bushell, E., Graham, A.L., Cameron, R.*, et al.* (2014). A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. Nature *507*, 253-257.

Smith, T.K., Vasileva, N., Gluenz, E., Terry, S., Portman, N., Kramer, S., Carrington, M., Michaeli, S., Gull, K., and Rudenko, G. (2009). Blocking variant surface glycoprotein synthesis in Trypanosoma brucei triggers a general arrest in translation initiation. PloS one *4*, e7532.

Soppa, U., and Becker, W. (2015). DYRK protein kinases. Curr Biol *25*, R488-489.

Spence, P.J., Brugat, T., and Langhorne, J. (2015). Mosquitoes Reset Malaria Parasites. PLoS Pathog *11*, e1004987.

Spence, P.J., Jarra, W., Levy, P., Reid, A.J., Chappell, L., Brugat, T., Sanders, M., Berriman, M., and Langhorne, J. (2013). Vector transmission regulates immune control of Plasmodium virulence. Nature *498*, 228-231.

Spoerri, I., Chadwick, R., Renggli, C.K., Matthews, K., Roditi, I., and Burkard, G. (2007). Role of the stage-regulated nucleoside transporter TbNT10 in differentiation and adenosine uptake in Trypanosoma brucei. Mol Biochem Parasitol *154*, 110-114.

Stanne, T.M., and Rudenko, G. (2010). Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. Eukaryot Cell *9*, 136-147.

Stephens, N.A., and Hajduk, S.L. (2011). Endosomal localization of the serum resistance-associated protein in African trypanosomes confers human infectivity. Eukaryot Cell *10*, 1023-1033.

Steverding, D., Stierhof, Y.D., Chaudhri, M., Ligtenberg, M., Schell, D., Beck-Sickinger, A.G., and Overath, P. (1994). ESAG 6 and 7 products of Trypanosoma brucei form a transferrin binding protein complex. European journal of cell biology *64*, 78-87.

Stijlemans, B., De Baetselier, P., Caljon, G., Van Den Abbeele, J., Van Ginderachter, J.A., and Magez, S. (2017). Nanobodies As Tools to Understand, Diagnose, and Treat African Trypanosomiasis. Frontiers in immunology *8*, 724.

Strickler, J.E., and Patton, C.L. (1975). Adenosine 3',5'-monophosphate in reproducing and differentiated trypanosomes. Science *190*, 1110-1112.

Sudarshi, D., Lawrence, S., Pickrell, W.O., Eligar, V., Walters, R., Quaderi, S., Walker, A., Capewell, P., Clucas, C., Vincent, A.*, et al.* (2014). Human African

trypanosomiasis presenting at least 29 years after infection--what can this teach us about the pathogenesis and control of this neglected tropical disease? PLoS Negl Trop Dis *8*, e3349.

Szoor, B., Dyer, N.A., Ruberto, I., Acosta-Serrano, A., and Matthews, K.R. (2013). Independent pathways can transduce the life-cycle differentiation signal in Trypanosoma brucei. PLoS Pathog *9*, e1003689.

Szoor, B., Ruberto, I., Burchmore, R., and Matthews, K.R. (2010). A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway. Genes Dev *24*, 1306-1316.

Szoor, B., Wilson, J., McElhinney, H., Tabernero, L., and Matthews, K.R. (2006). Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. J Cell Biol *175*, 293-303.

Terrao, M., Marucha, K.K., Mugo, E., Droll, D., Minia, I., Egler, F., Braun, J., and Clayton, C. (2018). The suppressive cap-binding complex factor 4EIP is required for normal differentiation. Nucleic Acids Res.

Torreele, E., Bourdin Trunz, B., Tweats, D., Kaiser, M., Brun, R., Mazue, G., Bray, M.A., and Pecoul, B. (2010). Fexinidazole--a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness. PLoS Negl Trop Dis *4*, e923.

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S.A., Pinto, A., Van Den Abbeele, J.*, et al.* (2016). Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. Cell Host Microbe *19*, 837-848.

Turner, C.M. (1990). The use of experimental artefacts in African trypanosome research. Parasitol Today *6*, 14-17.

Turner, C.M., and Barry, J.D. (1989). High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. Parasitology *99 Pt 1*, 67-75.

Turner, C.M., Barry, J.D., Maudlin, I., and Vickerman, K. (1988a). An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. Parasitology *97 (Pt 2)*, 269-276.

Turner, C.M., Barry, J.D., and Vickerman, K. (1988b). Loss of variable antigen during transformation of Trypanosoma brucei rhodesiense from bloodstream to procyclic forms in the tsetse fly. Parasitology research *74*, 507-511.

Turner, C.M.R. (1997). The rate of antigenic variation in fly-transmitted and syringe-passaged infections of Trypanosoma brucei. Fems Microbiology Letters *153*, 227-231.

Tyler, K.M., Matthews, K.R., and Gull, K. (1997). The bloodstream differentiation-division of Trypanosoma brucei studied using mitochondrial markers. Proc Biol Sci *264*, 1481-1490.

Uliana, S.R., Goyal, N., Freymuller, E., and Smith, D.F. (1999). Leishmania: overexpression and comparative structural analysis of the stage-regulated meta 1 gene. Experimental parasitology *92*, 183-191.

Ullmann, A., Jacob, F., and Monod, J. (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. Journal of molecular biology *24*, 339-343.

Ullu, E., Matthews, K.R., and Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Molecular and cellular biology *13*, 720-725.

Unciti-Broceta, J.D., Arias, J.L., Maceira, J., Soriano, M., Ortiz-Gonzalez, M., Hernandez-Quero, J., Munoz-Torres, M., de Koning, H.P., Magez, S., and Garcia-Salcedo, J.A. (2015). Specific Cell Targeting Therapy Bypasses Drug Resistance Mechanisms in African Trypanosomiasis. PLoS Pathog *11*, e1004942.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. Nucleic Acids Res *40*, e115.

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homble, F., Grelard, A., Zhendre, V., Nolan, D.P., Lins, L.*, et al.* (2013). Mechanism of Trypanosoma brucei gambiense resistance to human serum. Nature *501*, 430-434.

Van der Ploeg, L.H., Schwartz, D.C., Cantor, C.R., and Borst, P. (1984). Antigenic variation in Trypanosoma brucei analyzed by electrophoretic separation of chromosome-sized DNA molecules. Cell *37*, 77-84.

van Leeuwen, F., Wijsman, E.R., Kieft, R., van der Marel, G.A., van Boom, J.H., and Borst, P. (1997). Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. Genes Dev *11*, 3232-3241.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A.*, et al.* (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature *422*, 83-87.

Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Van Xong, H., and Pays, E. (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. Mol Microbiol *36*, 328-340.

Vanhollebeke, B., Uzureau, P., Monteyne, D., Perez-Morga, D., and Pays, E. (2010). Cellular and molecular remodeling of the endocytic pathway during differentiation of Trypanosoma brucei bloodstream forms. Eukaryot Cell *9*, 1272-1282.

Vasquez, J.J., Hon, C.C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res *42*, 3623-3637.

Vassella, E., Den Abbeele, J.V., Butikofer, P., Renggli, C.K., Furger, A., Brun, R., and Roditi, I. (2000). A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes Dev *14*, 615-626.

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997a). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J Cell Sci *110 (Pt 21)*, 2661-2671.

Vassella, E., Straesser, K., and Boshart, M. (1997b). A mitochondrion-specific dye for multicolour fluorescent imaging of Trypanosoma brucei. Mol Biochem Parasitol *90*, 381-385.

Vickerman, K. (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. Nature *208*, 762-766.

Vickerman, K. (1978). Antigenic variation in trypanosomes. Nature *273*, 613-617.

Wall, R.J., Rico, E., Lukac, I., Zuccotto, F., Elg, S., Gilbert, I.H., Freund, Y., Alley, M.R.K., Field, M.C., Wyllie, S.*, et al.* (2018). Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. Proc Natl Acad Sci U S A *115*, 9616-9621.

Walrad, P.B., Capewell, P., Fenn, K., and Matthews, K.R. (2012). The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in Trypanosoma brucei. Nucleic Acids Res *40*, 2869-2883.

Wamwiri, F.N., and Changasi, R.E. (2016). Tsetse Flies (Glossina) as Vectors of Human African Trypanosomiasis: A Review. BioMed research international *2016*, 6201350.

Watkins, J.F. (1964). OBSERVATIONS ON ANTIGENIC VARIATION IN A STRAIN OF TRYPANOSOMA BRUCEI GROWING IN MICE. The Journal of hygiene *62*, 69-80.

Welburn, S.C., Coleman, P.G., Maudlin, I., Fevre, E.M., Odiit, M., and Eisler, M.C. (2006). Crisis, what crisis? Control of Rhodesian sleeping sickness. Trends Parasitol *22*, 123-128.

Welply, J.K., Fowler, A.V., and Zabin, I. (1981). beta-Galactosidase alpha-complementation. Overlapping sequences. J Biol Chem *256*, 6804-6810.

WHO (2012). Accelerating work to overcome the global impact of neglected tropical diseases. A roadmap for implementation. (Geneva: World Health Organization).

WHO (2013). Control and surveillance of human African trypanosomiasis (Geneva: World Health Organization).

WHO (2018). Trypanosomiasis, human African (sleeping sickness). (World Health Organization).

Wickstead, B., Ersfeld, K., and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome Res *14*, 1014-1024.

Wiedemar, N., Graf, F.E., Zwyer, M., Ndomba, E., Kunz Renggli, C., Cal, M., Schmidt, R.S., Wenzler, T., and Maser, P. (2018). Beyond immune escape: a variant surface glycoprotein causes suramin resistance in Trypanosoma brucei. Mol Microbiol *107*, 57-67.

Wu, Y., Haghighat, N.G., and Ruben, L. (1992). The predominant calcimedins from Trypanosoma brucei comprise a family of flagellar EF-hand calcium-binding proteins. The Biochemical journal *287 (Pt 1)*, 187-193.

Wurst, M., Seliger, B., Jha, B.A., Klein, C., Queiroz, R., and Clayton, C. (2012). Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in Trypanosoma brucei. Mol Microbiol *83*, 1048-1063.

Xong, H.V., Vanhamme, L., Chamekh, M., Chimfwembe, C.E., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P., and Pays, E. (1998). A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. Cell *95*, 839-846.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. Protein & cell *2*, 202-214.

Yabu, Y., and Takayanagi, T. (1988). Trypsin-stimulated transformation of Trypanosoma brucei gambiense bloodstream forms to procyclic forms in vitro. Parasitology research *74*, 501-506.

Ziegelbauer, K., Quinten, M., Schwarz, H., Pearson, T.W., and Overath, P. (1990). Synchronous differentiation of Trypanosoma brucei from bloodstream to procyclic forms in vitro. European journal of biochemistry *192*, 373-378.

Zimmermann, H., Subota, I., Batram, C., Kramer, S., Janzen, C.J., Jones, N.G., and Engstler, M. (2017). A quorum sensing-independent path to stumpy development in Trypanosoma brucei. PLoS Pathog *13*, e1006324.

Appendix A: Antibody concentrations

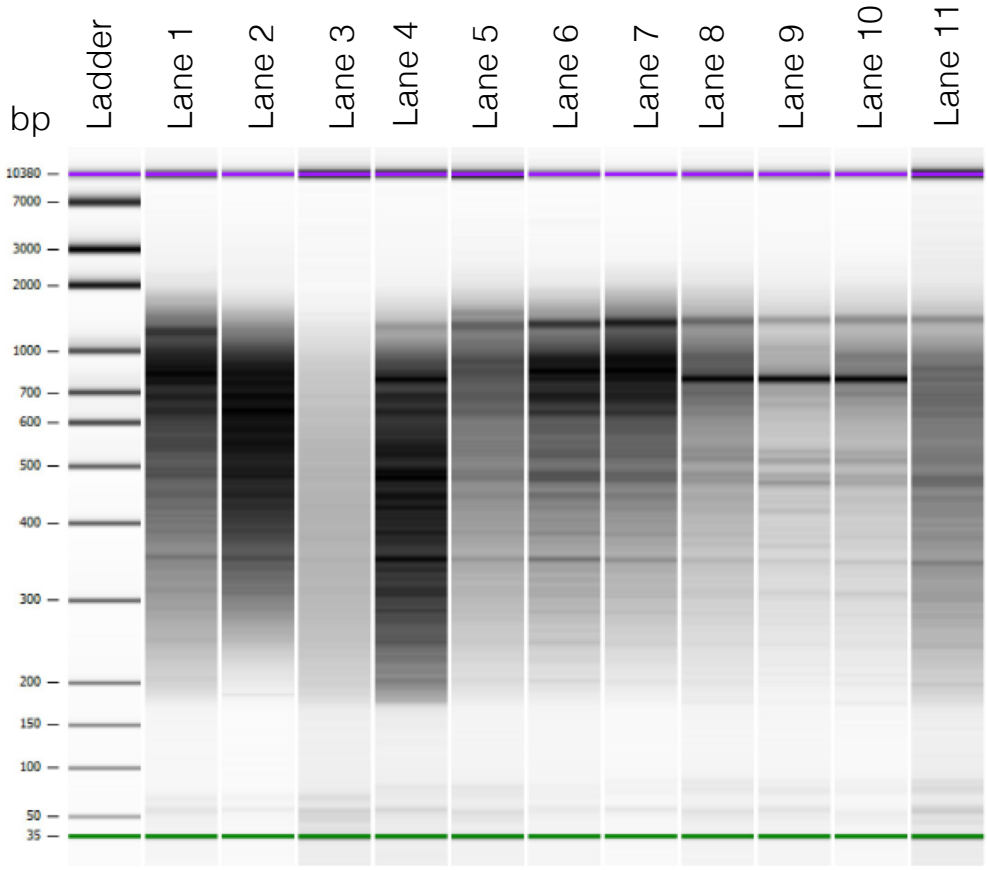
Table A: The working concentrations for the antibodies used in this study.

ANTIBODY	APPLICATION	CONCENTRATION
Rabbit α-VSG AnTat1.1 (Jay Bangs, USA)	FACS IFA	1:20,000 1:20,000
Rabbit α-PAD1 (Eurogentec)	IFA Western	1:1000 1:1000
Mouse α-EF1α (Millipore)	FACS Western	1:5000 1:7000
Alexa Fluor® 647-conjugated α-VSG 221 (Monica Mugnier, USA)	FACS	1:5000
α-rabbit Cy5 (Jackson ImmunoResearch)	FACS	1:1000
α-rabbit Alexa Fluor® 568 (Invitrogen)	IFA (PAD1) IFA (VSG AnTat1.1)	1:500 1:1000
α-rabbit IRDye® 800CW (LI-COR)	Western	1:7000
α-mouse IRDye® 560CW (LI-COR)	Western	1:5000

Appendix B: Bioanalyzer profiles of 11 VSGseq library preparations

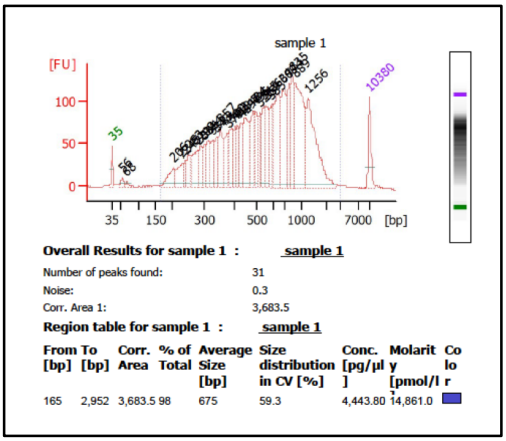
Libraries were analysed for efficient fragmentation during the tagmentation step by in-house providers on an Agilent 2100 Bioanalyzer at Johns Hopkins Bloomberg School of Public Health (USA). Typical libraries will consist of a range of fragments between 250-1000bp. Sample identities are provided above each profile in Figure B(2).

(1)



(2)

LANE 1: HYP2 +DOX3 PRE-MACS



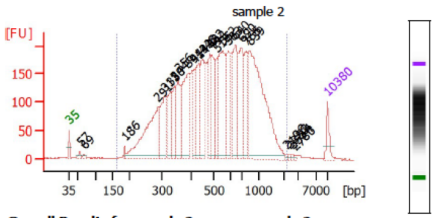
Overall Results for sample 1 : **sample 1**

Number of peaks found:	31
Noise:	0.3
Corr. Area 1:	3,683.5

Region table for sample 1 : **sample 1**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
165	2,952	3,683.5	98	675	59.3	4,443.80	14,861.0	

LANE 2: HYP2 –DOX1 POST-MACS



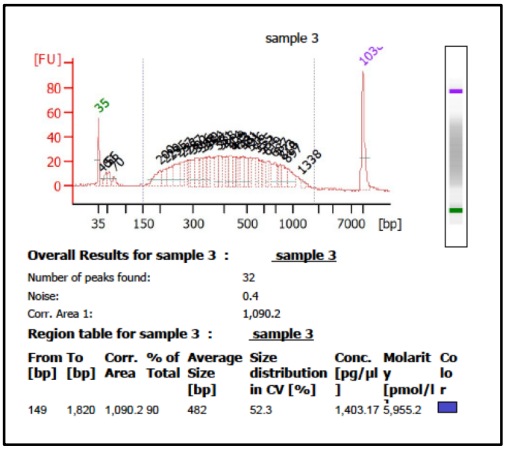
Overall Results for sample 2 : **sample 2**

Number of peaks found:	25
Noise:	0.3
Corr. Area 1:	6,435.1

Region table for sample 2 : **sample 2**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
164	2,170	6,435.1	98	596	51.4	7,718.07	26,096.5	

LANE 3: HYP2 –DOX2 POST-MACS



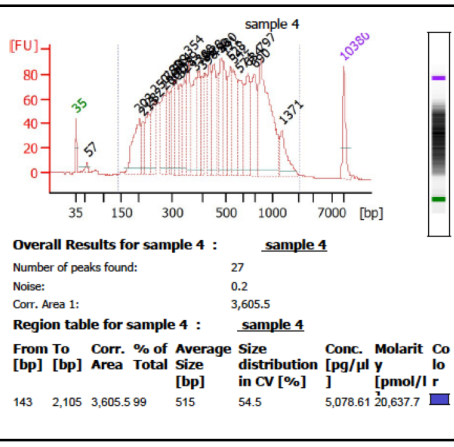
Overall Results for sample 3 : **sample 3**

Number of peaks found:	32
Noise:	0.4
Corr. Area 1:	1,090.2

Region table for sample 3 : **sample 3**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
149	1,820	1,090.2	90	482	52.3	1,403.17	5,955.2	

LANE 4: NEK -DOX1 PRE-MACS



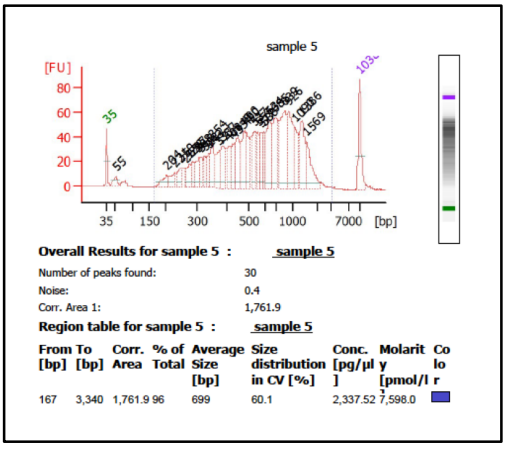
Overall Results for sample 4 : **sample 4**

Number of peaks found:	27
Noise:	0.2
Corr. Area 1:	3,605.5

Region table for sample 4 : **sample 4**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
143	2,105	3,605.5	99	515	54.5	5,078.61	20,637.7	

LANE 5: NEK –DOX2 PRE-MACS



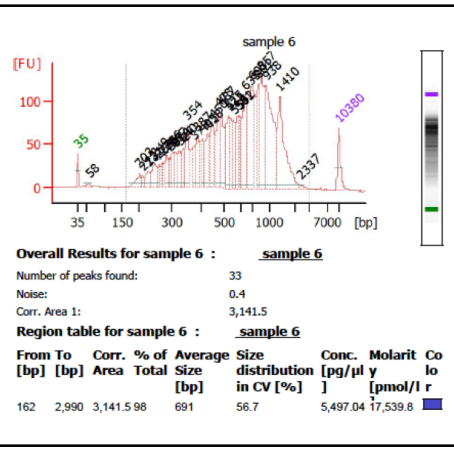
Overall Results for sample 5 : **sample 5**

Number of peaks found:	30
Noise:	0.4
Corr. Area 1:	1,761.9

Region table for sample 5 : **sample 5**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
167	3,340	1,761.9	96	699	60.1	2,337.52	7,598.0	

LANE 6: NEK -DOX3 PRE-MACS



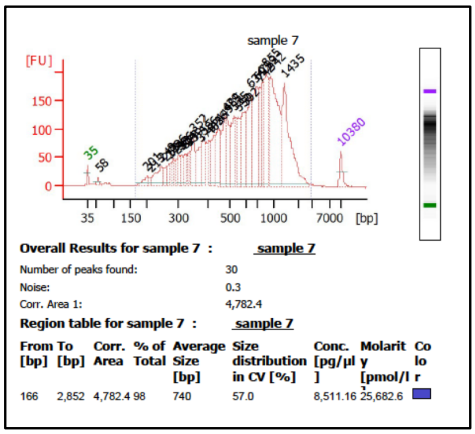
Overall Results for sample 6 : **sample 6**

Number of peaks found:	33
Noise:	0.4
Corr. Area 1:	3,141.5

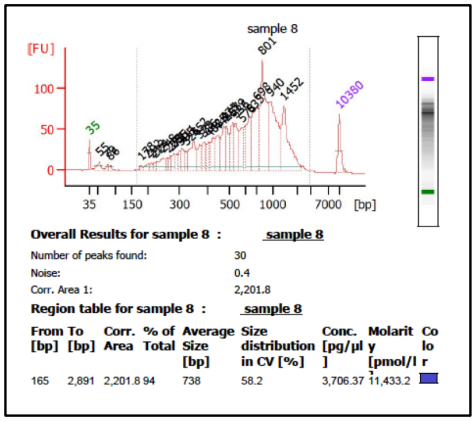
Region table for sample 6 : **sample 6**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
162	2,990	3,141.5	98	691	56.7	5,497.04	17,539.8	

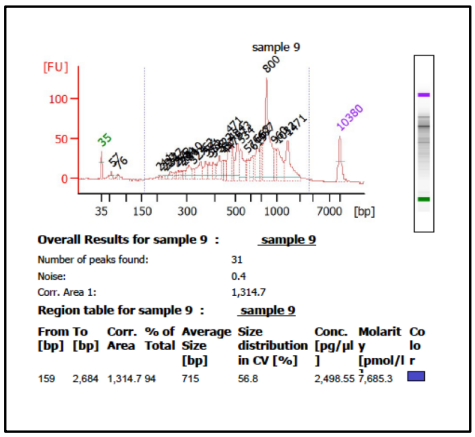
LANE 7: DYRK –DOX1 PRE-MACS



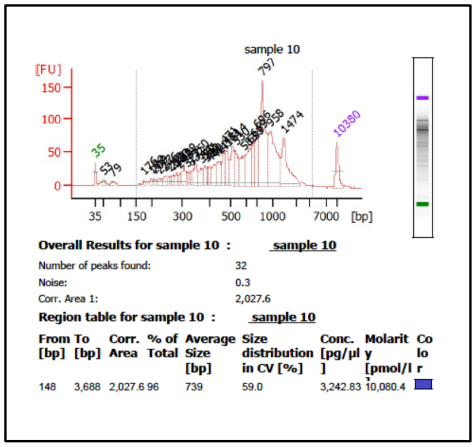
LANE 8: DYRK –DOX2 PRE-MACS



LANE 9: DYRK –DOX3 PRE-MACS



LANE 10: DYRK +DOX1 PRE-MACS



LANE 11: DYRK +DOX3 PRE-MACS

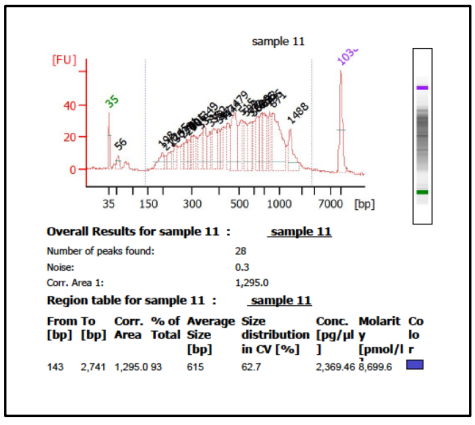


Figure B: The VSGseq library preparations were efficiently fragmented (1) 'Gel-like' image generated by the bioanalyzer after the electrophoretic separation of the DNA

fragments. The purple band represents the high molecular weight marker and the green band the low molecular weight marker **(2)** Fragment sizes plotted for each of the individual library preparation samples. The replicate name is stated above each plot.

Appendix C: Selected monomorphs downregulated transcripts

Table C: The shortlist of 68 transcripts which were significantly downregulated in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs compared to the parental pleomorphs, in order of most to least downregulated.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	NOTES	LOG2FC IF INCLUDED IN INT1 RELATIVE TO START DATASET
Tb927.1.5260	Hypothetical protein	-1.81	0.0040	GPI-anchor and 2 TM domains predicted.	-
Tb927.6.510	GPEET procyclin	-1.63	0.0003	Major GPI-anchored surface coat protein in procyclic forms (Bütikofer et al., 1997).	-1.39
Tb927.5.4020	Hypothetical protein	-1.50	0.0005	Member of cell surface phylome Fam50, clade III (Jackson et al., 2013). No transmembrane domain/GPI anchor predicted but has a signal peptide. Stabilised by ZC3H20 (Ling et al., 2011). Protein has BLASTP similarity to TolA, a bacterial surface transporter.	-1.33
Tb927.7.2660	ZC3H20	-1.41	0.0004	Zinc finger protein. Required for growth of procyclic forms. Binds and stabilises at least two developmentally regulated procyclic specific RNAs (Ling et al., 2011).	-1.59
Tb927.8.8320	Hypothetical protein, conserved	-1.26	0.0012	-	-1.15
Tb927.7.5770	PK53	-1.20	0.0079	Nuclear Dbf2 related kinase. Knockdown leads to defect in cytokenesis and kinetoplast duplication/segregation (Jones et al., 2014).	-0.92
Tb927.6.810	Hypothetical protein	-1.18	0.0229	-	-
Tb927.9.7470	TbNT10	-1.06	0.0003	Purine nucleoside transporter (Spoerri et al., 2007).	-1.18
Tb927.11.15010	NEK21	-1.00	0.0029	Serine/threonine-protein kinase. Bound by RBP10 (Mugo and Clayton, 2017).	-1.16
Tb927.7.5940	Protein Associated with Differentiation 2 (PAD2)	-0.97	0.0003	Recieves CCA procyclic differentiation signal (Dean et al., 2009).	-0.93
Tb927.10.10260	EP1 procyclin	-0.96	0.0056	Major GPI-anchored surface coat protein in procyclic forms (Bütikofer et al., 1997).	-0.61
Tb927.5.2260	Hypothetical protein, conserved	-0.93	0.0015	Bound by RBP10 (Mugo and Clayton, 2017). META domain.	-
Tb927.9.4560	RBP32	-0.92	0.0004	Bound by RBP10 (Mugo and Clayton, 2017).	-0.91
Tb927.10.12780	ZC3H37	-0.91	0.0056	Zinc finger protein. Increased blasticidin resistance when tethered to a drug resistance marker (Erben et al., 2014).	-1.15
Tb927.7.6330	Hypothetical protein	-0.90	0.0020	Bound by RBP10 (Mugo and Clayton, 2017). 4 TM domains predicted.	-0.79
Tb927.11.6890	TbPIF1	-0.90	0.0136	DNA repair and recombination helicase protein. Essential for kinetoplast minicircle replication (Liu et al., 2010). Bound by RBP10 (Mugo and Clayton, 2017).	-1.00
Tb927.10.8730	hypothetical protein, conserved	-0.90	0.0036	ABC1 family.	-0.94
Tb927.7.6010	Hypothetical protein	-0.89	0.0492	-	-
Tb927.10.8740	Hypothetical protein	-0.88	0.0449	1 TM domain and signal peptide predicted.	-
Tb11.v5.0513	Hypothetical protein, conserved	-0.83	0.0354	NLI interacting factor-like phosphatase domain	-0.86
Tb927.10.15620	Hypothetical protein	-0.82	0.0091	3 TM domains and signal peptide predcited.	-0.73
Tb927.5.810	ZC3H11	-0.82	0.0154	Zinc finger protein. Binds chaperones required for heat shock response during procyclic differentiation (Droll et al., 2013)	-0.90
Tb927.6.860	Hypothetical protein	-0.81	0.0033	Bound by RBP10 (Mugo and Clayton, 2017).	-0.91
Tb927.3.1650	Hypothetical protein	-0.81	0.0422	Bound by RBP10 (Mugo and Clayton, 2017). 2 TMs predicted.	-

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	NOTES	LOG2FC IF INCLUDED IN INT1 RELATIVE TO START DATASET
Tb927.11.16820	Hypothetical protein, conserved	-0.80	0.0395	Signal peptide predicted.	-
Tb927.11.2410	Hypothetical protein, conserved	-0.80	0.0045	Bound by RBP10 (Mugo and Clayton, 2017). BAR/IMD domain-like.	-0.79
Tb927.11.8340	Hypothetical protein	-0.79	0.0227	3 TMs and signal peptide predicted.	-
Tb927.11.810	Hypothetical protein, conserved	-0.79	0.0431	TPR-like domain.	-
Tb927.4.1000	Hypothetical protein, conserved	-0.79	0.0017	Bound by RBP10 (Mugo and Clayton, 2017).	-1.05
Tb927.9.11050	4E-interacting protein (4EIP)	-0.78	0.0017	Contributes to tranlsation repression during differentiation to stumpy form (Terrao et al., 2018).	-0.59
Tb927.10.3360	Hypothetical protein, conserved	-0.77	0.0011	Active form triggers differentiation to procyclic forms. Glycosomal enzyme (Szoor et al., 2010)	-0.81
Tb927.10.8050	TFIIF-stimulated CTD (C terminal domain) phosphatase, putative	-0.75	0.0064	In yeast and mammals, TFIIF-stimulated CTD phosphatase (FCP1) catalyses the dephosphorylation of the CTD of RNApolII (Kobor et al., 1999). Bound by RBP10 (Mugo and Clayton, 2017).	-0.91
Tb927.10.10770	Generative cell specific 1 (GCS1) protein	-0.75	0.0009	In plants, GCS1 localises to PM of generative cells. Essential fertilisation factor in angiosperms, required for guidance of sperm cells to ovules (Mori et al., 2006; von Besser et al., 2006). Bound by RBP10 (Mugo and Clayton, 2017).	-0.66
Tb11.1390	Hypothetical protein, conserved	-0.74	0.0227	6 TM domains and signal peptide predicted.	-0.60
Tb927.5.2160	Hypothetical protein, conserved	-0.74	0.0057	Stabilised by ZC3H20 (Ling et al., 2011). META domain.	-0.59
Tb927.7.6300	Hypothetical protein, conserved	-0.72	0.0126	Bound by RBP10 (Mugo and Clayton, 2017). WD40 domain.	-0.62
Tb927.3.1640	Hypothetical protein	-0.71	0.0056	3 TM domains predicted.	-0.67
Tb927.11.240	Hypothetical protein, conserved	-0.71	0.0012	Signal peptide predicted.	-0.49
Tb927.10.5100	Hypothetical protein	-0.71	0.0133	3 TM domains predicted.	-0.50
Tb927.9.2320	Hypothetical protein, conserved	-0.71	0.0016	Bound by RBP10 (Mugo and Clayton, 2017). Methyltransferase domain.	-0.60
Tb927.8.4300	Hypothetical protein	-0.70	0.0306	Bound by RBP10 (Mugo and Clayton, 2017). 3 TM domains predicted.	-
Tb927.10.12060	Hypothetical protein, conserved	-0.70	0.0069	Bound by RBP10 (Mugo and Clayton, 2017). F-box domain.	-0.61
Tb927.8.5350	Hypothetical protein, conserved	-0.69	0.0069	Bound by RBP10 (Mugo and Clayton, 2017).	-0.59
Tb927.3.5760	Hypothetical protein, conserved	-0.69	0.0009	Signal peptide predicted.	-0.81
Tb927.11.6040	Hypothetical protein, conserved	-0.67	0.0029	MFS general substrate transporter family	-0.76
Tb927.7.5930	Protein Associated with Differentiation 1 (PAD1)	-0.67	0.0014	Expressed on surface of stumpy forms (Dean et al., 2009).	-0.59
Tb927.11.11560	DNA topoisomerase II, putative	-0.66	0.0003	Nuclear. Role in trypanosomes unclear (Kulikowicz and Shapiro, 2006).	-0.75
Tb927.3.2920	Hypothetical protein, conserved	-0.66	0.0126	Bound by RBP10 (Mugo and Clayton, 2017).	-0.59
Tb927.8.5870	Hypothetical protein, conserved	-0.65	0.0026	-	-0.54
Tb927.7.520	Hypothetical protein, conserved	-0.65	0.0075	Bound by RBP10 (Mugo and Clayton, 2017). Cytochrome b5-like Heme/Steroid binding domain.	-0.59
Tb927.11.14070	Repressor of differentiation kinase 1 (RDK1)	-0.65	0.0023	Negative regulator of procyclic differentiation (Jones et al., 2014). Bound by RBP10 (Mugo and Clayton, 2017).	-0.74
Tb927.6.3880	Hypothetical protein, conserved	-0.64	0.0274	Bound by RBP10 (Mugo and Clayton, 2017).	-0.72
Tb927.10.3700	Gamma regulatory subunit of AMP-activated protein kinase (AMPK)	-0.64	0.0230	RNAi in PCF forms induces changes in surface molecule expression. Express both GPEET and EP1 (Clemmens et al., 2009).	-0.55

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	NOTES	LOG2FC IF INCLUDED IN INT1 RELATIVE TO START DATASET
Tb927.4.2410	Hypothetical protein, conserved	-0.64	0.0017	Bound by RBP10 (Mugo and Clayton, 2017). Glycosyl hydrolase family 65 central catalytic domain.	-0.73
Tb927.5.320	Receptor-type adenylate cyclase (GRESAG 4)	-0.64	0.0024	Midgut specific adenyl cyclase, marker for late procyclic stage (Imhof et al., 2014). Bound by RBP10 (Mugo and Clayton, 2017).	-0.57
Tb11.v5.0209	Hypothetical protein, conserved	-0.64	0.0043	Adenylate and Guanylate cyclase catalytic domain. Signal peptide predicted.	-0.56
Tb927.11.2690	Succinyl-CoA:3-ketoacid coenzyme A transferase	-0.64	0.0229	Mitochondrial enzyme (Vertommen et al., 2008).	-0.80
Tb927.7.6830	Trans-sialidase (putative)	-0.64	0.0006	Adenylate and Guanylate cyclase catalytic domain. Signal peptide predicted.	-0.50
Tb927.9.5900	Glutamate dehydrogenase	-0.63	0.0029	Mitochondrial enzyme (Mantilla et al., 2017).	-0.67
Tb927.11.2450	Hypothetical protein	-0.62	0.0092	Bound by RBP10 (Mugo and Clayton, 2017).	-
Tb927.11.2480	Hypothetical protein	-0.62	0.0328	2 TMs predicted.	-
Tb927.5.285b	ESAG4	-0.62	0.0031	Modulation of the host innate immune response (Salmon et al., 2012).	-0.55
Tb11.v5.0217	Receptor-type adenylate cyclase (GRESAG 4)	-0.62	0.0037	Midgut specific adenyl cyclase, marker for late procyclic stage (Imhof et al., 2014). Stabilised by ZC3H20 (Ling et al., 2011).	-0.57
Tb927.1.2200	Zinc finger protein family member, putative	-0.62	0.0302	-	-
Tb927.9.7540	Calpain-like cysteine peptidase	-0.62	0.0024	Calcium-dependant activity. Bound by RBP10 (Mugo and Clayton, 2017).	-0.62
Tb927.10.15610	Zinc finger protein, putative	-0.62	0.0299	Bound by RBP10 (Mugo and Clayton, 2017).	-0.73
Tb927.11.1830	Hypothetical protein, conserved	-0.61	0.0004	6 TMs predicted. Bound by RBP10 (Mugo and Clayton, 2017).	-0.51
Tb927.11.13940	Ubiquitin-conjugating enzyme E2, putative	-0.61	0.0041	E2 one of three enzymes (E1-3) which bring about polyubiquitination (Hershko et al., 1983).	-0.61

Appendix D: Selected monomorphs upregulated transcripts

Table D: The shortlist of 11 transcripts which were significantly upregulated in the GFP^{ESpro}AnTat1.1^{ES} *NEK* RNAi selected monomorphs compared to the parental pleomorphs, in order of most to least upregulated.

GENE CODE	ANNOTATION	LOG2 FC	ADJUSTED P VALUE	NOTES	LOG2FC IF INCLUDED IN INT1 RELATIVE TO START DATASET
Tb927.7.3250	ESAG 6	0.73	0.02766	Transferrin receptor (Steverding et al., 1994).	0.64
Tb927.7.3260	ESAG 7	0.70	0.02192	Transferrin receptor (Steverding et al., 1994).	0.61
Tb927.9.11580	Gim5a	0.59	0.00615	One monomer of the glycosomal membrane protein dimer, GIM5 (Maier et al., 2001).	0.38
Tb927.4.4860	AAT8, putative	0.58	0.00003	Amino acid transporter. Downregulated in Repressor of Differentiation Kinase 1 (RDK1) RNAi screen (Jones et al., 2014).	0.41
Tb927.9.1960	Nitrilase (NIT1), putative	0.57	0.00001	Metabolic repair enzyme. Catalyses the hydrolysis of deaminated glutathione (Peracchi et al., 2017).	0.47
Tb927.11.7700	Hypothetical protein	0.57	0.00020	Potential ortholog of p27 (mitochondrial membrane protein in Leishmania).	0.37
Tb927.8.5460	44 kDa calcimedin, Tb44	0.52	0.00002	Calcium-binding protein which localises to the FP. In vivo knockdown associated with prologation of infection and increased survival of mice (Emmer et al., 2010).	0.37
Tb927.4.2310	Asparaginyl-tRNA synthetase, putative	0.51	0.00000	Charge tRNAs with asparagine. Asparagine plays key role in biosynthesis of glycoproteins.	0.21
Tb927.11.5520	Triosephosphate isomerase (TPI or TIM)	0.51	0.00661	Glycosomal enzme (Swinkels et al., 1986).	-
Tb927.6.4280	Glyceraldehyde 3-phosphate dehydrogenase	0.50	0.00333	Glycosomal enzme (Misset et al., 1987).	0.43
Tb927.6.4300	Glyceraldehyde 3-phosphate dehydrogenase	0.50	0.00302	Glycosomal enzme (Misset et al., 1987).	0.42

Appendix E: SNPs identified in the selected monomorphs' genome

Table E: The 82 SNPs which occurred in the assembled selected monomorph genome compared to the parental pleomorph genome. The SNPs have been filtered to only include those that fall within the coding sequence of a gene which was identified as being significantly upregulated or downregulated in the selected monomorphs population compared to the parental pleomorph.

GENE	ANNOTATION	LOG2FC	SNP
Tb927.5.4020	Hypothetical protein	-1.5	A_G G_A G_A
Tb927.7.2660	ZC3H20	-1.41	T_C C_T
Tb927.8.8320	Hypothetical protein	-1.26	T_G C_T T_C
Tb927.7.5770	PK53	-1.2	C_T C_T T_C T_C T_C A_G G_T
Tb927.6.810	Hypothetical protein	-1.18	T_C
Tb927.7.5940	PAD2	-0.97	C_T A_C
Tb927.10.12780	ZC3H37	-0.91	T_C GT_G G_C
Tb927.11.6890	TbPIF1	-0.9	C_G T_G G_A A_G G_A
Tb11.v5.0513	Hypothetical protein	-0.83	G_A
Tb927.5.810	ZC3H11	-0.82	A_G
Tb927.6.860	Hypothetical protein	-0.81	T_G
Tb927.11.16820	Hypothetical protein	-0.8	A_G
Tb927.11.2410	Hypothetical protein	-0.8	G_C

GENE	ANNOTATION	LOG2FC	SNP
Tb927.11.810	Hypothetical Protein	-0.8	C_T G_T T_C G_A T_C T_C
Tb927.9.11050	4EIP	-0.78	G_C T_C T_G
Tb11.1390	Hypothetical Protein	-0.74	C_A C_G C_T G_A
Tb927.9.2320	Hypothetical protein	-0.71	C_T C_T C_T T_C
Tb927.11.6040	Hypothetical protein	-0.67	G_A
Tb927.7.5930	PAD1	-0.67	A_C
Tb927.11.11560	DNA Topoisomerase II	-0.66	A_G A_C G_A C_T A_G C_T G_A
Tb927.11.14070	RDK1	-0.65	C_T C_A C_G G_T A_G
Tb927.10.3700	Gamma regulatory subunit of AMPK	-0.64	G_A
Tb927.11.2690	Succinyl-CoA:3-ketoacid coenzyme A transferase	-0.64	T_C A_C
Tb927.7.6830	Trans-sialidase	-0.64	G_A C_T G_T A_G A_T
Tb11.v5.0217	GRESAG4	-0.62	G_A A_G G_A

GENE	ANNOTATION	LOG2FC	SNP
Tb927.10.15610	Putatative Zinc finger protein	-0.62	A_T G_A G_A C_T C_G A_G C_T C_A C_T
Tb927.4.4860	AAT8	0.58	C_T T_C
Tb927.9.1960	NIT1	0.57	A_G T_G A_G C_G
Tb927.4.2310	Asparaginyl-tRNA synthetase	0.51	T_G